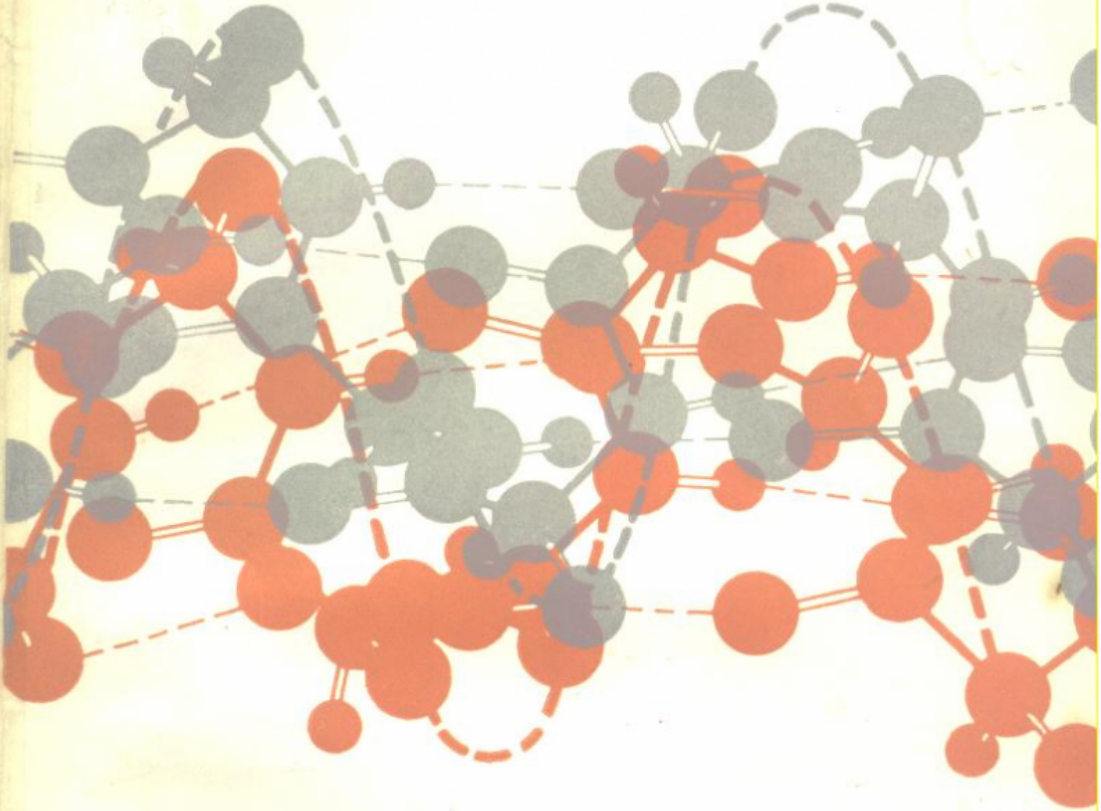


分子遗传学概论

遗传与进化中的信息流

〔美〕DowO．伍德沃德　VaIW．伍德沃德　著

葛扣麟　邹高治　倪大洲　李美华　杨金水　杨连福　译

上海科学技术出版社

分子遗传学概论

遗传与进化中的信息流

〔美〕 Dow O. 伍德沃德
Val W. 伍德沃德 著

葛扣麟 邹高治 倪大洲
李美华 杨金水 杨连福 译

上海科学技术出版社

CONCEPTS OF MOLECUI AR GENETICS

INFORMATION FLOW IN GENETICS AND EVOLUTION

DOW O. WOODWARD VAL W. WOODWARD

McGRAW-HILL BOOK COMPANY (1977)

分子遗传学概论

遗传与进化中的信息流

葛扣麟等译

上海科学技术出版社出版

(上海瑞金二路450号)

新华书店上海发行所发行　江苏扬中印刷厂印刷

开本 850×1156　1/32　印张 13.875　字数 367,000

1983年6月第1版　1983年6月第1次印刷

印数 1—10,600

统一书号：13119·1067 定价：(科五) 1.90元

前　言

究竟什么是严肃认真的科学活动，什么是不严肃认真的科学活动，在这个问题上，科学家之间有一个普遍的默契。绝大多数的科学家都同意这一约定俗成的指导法规，尽管目前还不是所有的人都无条件地赞助这一点。比如，人们不会反对，任何一个实验在其达到高度内在一致性之前都必须进行重复实践。人们同样赞成，实验中提出的科学结论不宜超乎实践所得资料的一定范围之外。还有，在必要时候要勇于把成果和荣誉让给他人，及时发表论文，不必过多的切磋琢磨，以致延误时机等等。

然而，在教育领域，教师们对于什么是切实的教授方式，什么是不合宜的教学活动，观点上便不尽一致了。虽然不能说是人各一辞，但究竟什么是构成优异教学法的结构因素，人们迄今还缺少相近的立场。从高等院校的教学角度来看，同样亦存在深刻的歧见。举例来说，关于教学的主要目标，是要求从全体学生中精选少数可望将来成为科学家的拔尖人材呢，还是着眼于提高全体学生对自然科学的理解能力。由于对教学目标的此种相互对立的看法，结果必然会相应在教学方式、价值标准、题材内容诸方面同样带来极其深刻的分歧。

教学体制本身构成了一整套按其广度和复杂性来说决不亚于其他学科的研究领域。可是一般从事于这一领域进行研究和工作的专业人员与那些沉缅于教学第一线的教师往往各唱各调，不通来往。这种情况对于那些专心致志讲授自然科学的学者更是如此。因为后者关心的仅是科学知识传授，他们在自然科学的教学活动中常常抱着某种想当然的观点，不恰当地把讲义内容与课堂效果等量齐观。结果，有些能够死记硬背讲义和教科书的学生反

而会博得这些先生们的青睐和赞誉。

现在科技界中已经制定出一些办法，以防止免受低劣科学的侵蚀(这里说的低劣并非指缺乏创见、平庸无味而言)。如果有人在实验中弄虚作假，渎职妄为，最终亦会被人识破，并查明其伪误所在。可是在涉及教学质量时，要是某个教师自比伯乐，把一些死读书而实际资质平平的学生当作宝贝、举荐于世，而另外一些教师则在提高全体学生对科学的理解力方面取得成绩，这里面的是非曲直便不大可能象前面所说，用一个试验或通过评论来求得澄清。诚然这类问题不属于自然科学研究的范畴，而是涉及到制度、政策的一类社会问题，它的解决需要有效地建立一套作为判断衡量的特定的社会价值标准。基于这一原因，应该说，一般的教师或学院机构对此无能为力。

Pablo Freire 还进一步举例说明了有关教育标准莫衷一是的困境。作者把教育描述为一种"感知、分析与发现"的过程。据他看来，把学生分列为不同智力等级的教学方法与教学计划根本称不上其为教育。Freire 进而提出，教育的目的就在于学习"对客观现实进行深入理解的过程"。于是我们又再次看到了另一种意见分歧：大多数西方科学家的实践活动可以追溯到十九世纪由逻辑实证论者所开创的传统。逻辑实证论者曾不太准确地告诉人们这一信条：只有借助于客观的科学资料，人们才有可能逐步接近于客观真理；与之相反，Freire 所指的客观现实却包含着相当程度的人为倾向性、社会价值标准、人类本性等有关因素，以及对人类未来设计中的主观意志论等等。显然这是完全不同的两种"现实"观。目前绝大多数科学家比较一致的看法是，科学资料本身应该与倾向性、价值标准、人类本性等两相无涉。但这并不意味着科学不具有目的性，或是科学有它独特的超自然属性，即在发展有关人类本性的理论中，存在有所谓的中性科学。

Freire 提出，教育必须寻找最大限度地使人类天赋的才智得以充分发展的手段和方法。其论点基于这一假设：人人都具有善于学习认识客观现实的能力。但千真万确的是，他所提出论点的

前提也和其他无数的巨大社会性前提一样并未从客观资料取得充分支持。而作为一个杰出的理论，取得客观资料的支持则是理所当然的事情。此点姑且不论，Freire 的论点还为风行某种不切实际的教学模式大开绿灯。如果他的论点能够科学地予以检验，那我们将可以由它来决定教学法的最佳模式。可是因为实际上做不到这点，结果大部分还是只能按传统方式来规定教学法的特点。当然，采取上述两种极端看法中间立场的，亦不乏其人。

从某些方面来说，自然科学可以十足有效，也可以毫无活力。那些对学习目的性高度明确的学生往往可以成为优秀的科学家，其中不少学生都曾得到过他们老师的启蒙和鼓励。而将来一心要成为一个伟大科学家的远大理想，对他们更无疑是巨大的精神动力(我们作为教师恐怕都有这种皮格马利翁式的现实体会*)。对这类学生所进行的自然科学教育便能转变为将来求职的有利因素。可是就在这些具有高度学习目的性的学生顺利走向专业工作的同时，大多数学生却在科学进军的道路上抱恨落伍了。他们不仅学不到专业科学技术，而且也掌握不了一般的基础科学，科学知识对一个生活在现代科技社会的公民可能得以大显身手的一切生活道路，实际上已被全部堵死。

爱因斯坦说过："除非是出现奇迹，否则现在的教学方法不可能不全部窒息掉人们基于好奇心而舒发的神圣探索精神，因为作为此种崇高"精神"的重要支柱之一的自由探索，其重要性与紧迫性决不亚于任何一种社会刺激。"不少教师曾以谋求"培养学生的此种好奇心和探索精神"为己任。可是面对着社会上实用主义的潮流、习惯势力以及教学上的行政措施等等，这不是一件轻而易举的事情，因为所有这些因素都构成了教学活动所必须所命的舞台，它规定着教学活动应该在什么时候，什么地方，以及以什么方式合宜适度地进行。

上面虽然只列举了教学方面的很少一部分问题（只能算是海

---

* 皮格马利翁(Pygmalion)，希腊神话中的塞浦路斯国王，热爱自己所雕塑的少女石像。——译注

滩一贝），读者必然会问，本书多少是否会有助于一个或几个问题以某种程度的解决？我们可以干脆利落地回答："即使有所裨益，亦必微不足道。"由于教学体制的复杂性，解决这一问题已远远超越本书，乃至一所大学力所能及的范围。但话虽如此，书籍、教师和高等院校在这方面的作用还是不容忽视的。如果我们把教育事业的一般状况看成仅仅是巨大社会的一个缩影，那末，不言而喻，要想通过书籍和教师的推动而使之可能产生某种变化当然不会很快。事实上，任何一本书的潜在作用整个地都仰仗于教师的力量，本书亦概莫能外。如果教师谋求要在自然科学的教学工作中赋予某种新意，本书或许能助一臂之力，否则，它也只能以传统习惯的方式作为一本普通的教科书使用而已。

据我们看来，本书在下列几方面或许对读者有些帮助。第一，作为科学家，人们将可由之享受到某种科学进展带来的乐趣和激情。我们将力求在书中多多介绍这方面的资料和科学研究的方法，以提供给那些尚未亲自参加过科学实践的青年们作参考。因为不亲身经历科学舞台的具体实践，决不会懂得什么是演员、什么是剧情、什么是舞台道具。要是这方面的知识能综述得令人满意，对那些未曾身历其境的人来说将可借以领略一下科学研究中的无限风光，关于科学研究中至关重要的思维活动的特色也可作一番具体深入的体察。

第二，我们将试图表明关于遗传学、化学和进化论这些学科之间的关联性、连续性和不可分割性。即便上述每门学科涉猎不广，但我们亦谋求将之归入一个总的题目。这个总题目便叫做生物学中的信息流（information flow，信息传递）。从分子遗传学的概念来看，这种提法也是完全可以理解的。在书中将努力：（1）通过提出问题以及引导对这类问题的深入思考，以开创某些"认识事物的新途径"；（2）引入足够多的有关分子遗传学研究的实验，向读者表明，人们在研究一个充满偶然性的客观世界时，在成功解决问题方面，必然性与偶然性之间往往存在着某种微妙的平衡。（3）指出在分子遗传学的题材内容与其他自然科学学科和人类关心的问

题之间也有着密切的关系。为此我们力求使本书的内容能够适应于已具有一定遗传学与化学基础阅读水平和理解水平的学生。

如果能对科学家"思维活动"中的多种思路作一番身临其境的体察，将能形象生动地说明，一个人善于从失败和成功中汲取经验教训对于科学活动的有效进展是多么重要。通常，人们从数据和事实即得以形成概念和理论，进而由理论再上升为假说。反过来，一种假说又需要用实践再检验，通过实验又可获得更多的资料和数据。在如此往复循环，复杂多变的认识过程中，失败与成功的可能性随处可遇。我们在书中拟将各种成功与失败的典型实验都略举数端。读者可以看到，一些概念确立了，发展了，另一些概念又被否定了、抛弃了。但无论如何，即便本书对自然科学的讲授有微小贡献，这也必定仰赖于讲课教师的大力配合与辛勤劳动。

在本书的设计安排上，同样亦力求能提供多方面的途径，帮助学生去接触分子遗传学，特别是书中引述了为数甚多的具体例证。实验引出的结论并不一定均有可靠的证据。我们以图表、数据与模型的方法列出这类材料，便能提供给那些希望评论这类实验事例的学生以必要的第一手资料。学生们在这方面必然会发挥其有益的反馈相长作用。可以断言，他们肯定是会支持如此处理的。其中既包括抱雄心大志诚心想成为科学家的学生，也包含有在学习工作中吹毛求疵，穷于究竟的学生，甚至还有那些目前只是"随大流"的学生在内。

我们在征求对本书的意见中饶有兴味的是下面一些观点分歧。学生们普遍要求能多多列入有关科学、历史、哲学、社会危机等相互关系的知识；而科学家则又一致要求希望削减乃至删除这一类型的资料。我们在处理这一矛盾时，基本上是采取左右并蓄的折衷主义办法。

对于另外一些目标，我们感到即便不是绝无可能，但亦难以如愿以偿。例如我们写进分子遗传学一书的内容能否使其中每一章彼此在分量上恰如其分地合乎如图 1 所示的轻重关系和整体布局还是一个问题。另一方面，作为一本教科书，其写作的平铺直叙方

式，亦即书籍写作上的单向性与自然科学所固有的二维性甚至三维性的立体格局，在表达时很难对偶接笋。这里我们可再一次看到，对于教师来说，“表达”自然科学思维形态(二度空间与三度空间结构)的能力往往是因人而异的。理由十分简单，主要是太抽象难教了。

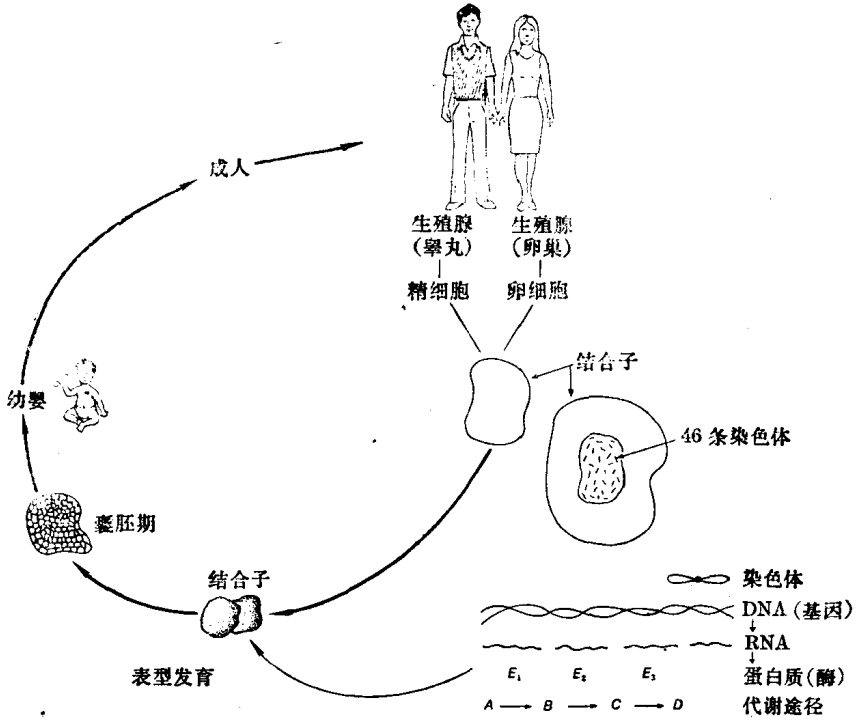


图 1　世代的纽带

即便生生死死的问题家喻户晓，但是人们却未必体会到怀胎和新生儿生长发育以及代谢过程的复杂性，上代与子代的关系问题一般人恐怕也是茫然无知的。上述简图可以进一步展开，如图右下角所示，这里还包含有更多的细节详情。

科学研究的方法在教学实践中常常是不予提及的，这样做的不良后果是留给学生的只是一大堆事实和数据。而对于至为重要的科学思维方法则等于不学(图 2)。

如果人们不考虑自然科学的研究方法，那么有关的快乐与烦

恼、沮丧与创新乃至自然科学的思想体系也就不再存在了。可是实际上无论是基础科学，还是专业科学两方面都离不开自然科学思想体系的指导。不少有丰富经验的科学家，他们都曾亲自努力从事于直观事物的研究，并有善于运用归纳法、演绎法去发现新事物的切身经历，因而有条件与学生共同去分享他们既有的经验，可是却常迫于传统势力的压力，只能跟着"逻辑式"的教学大纲亦步亦趋，从而最终放弃自己本可大施才华的一整套教学本领。这种情况与社会上的一股"反科学"恶浪有关，同时与对讲授自然科学的非难亦大有关系。这些指责往往来自于一些科学观念浅薄的学生，在他们看来，讲授自然科学不过是"求职前不得不由的说教"。

由于科学知识的累积增长愈来愈快，因之为学生挑选适当的课外知识这一点需要继续作为一个问题引起重视。但是究竟该如何掌握进行，解决这一任务的部分答案是在于要求科学家们能充分发挥他们的能动作用。把"科学奥秘"用浅显易懂的语言向人们

图 2　教室中的教学实践与实验中的科学实践

在课堂上听科学讲授的学生不会知道科学历史发展的全过程。一般地说，教师不会讲授科学史中有关"人的思维活动"的那些部分。有时，有关科学实验的那些部分在安排上也是按发展过程中事物本身的逻辑顺序来讲授的。

作普及介绍。另一方面则是为学生尽量撰写有关的综合性论文和评论。《科学美国人》(Scientific American)杂志便是一份能合宜提供版面的优秀读物。不少评论本书的人士指出，在科学家们中间有一种普遍的情绪，他们不乐意去干这一类事情。还有不少专业人员对于发挥他们的特点同样亦持相对保守和消极的态度。一种顾忌是害怕学生会因此而寻找逃避艰苦学习的"捷径"，另一种顾虑，则认为放弃老规矩，而去钻学习自然科学的新门径不合一个学者应有的气派。既然我们并不同意上面的观点，于是便理所当然地受到他们的批评。我们的看法是，作为教师应有义务尽可能地缩短学生们在接触科学知识伊始，到从中体会到身历其境的乐趣，以及到对科学技术产生深刻感受这一段时间的经历和距离。

无论如何，在今天强求教师如此照办，是既脱离实际，而又不公道的。因为大大扩增了的班级规模以及由此而大大加重了的教学负担已成为目前关系到教学质量的两个基本要素。而解决这些矛盾已经远远超出本书与教师的职责之外。面对着可能进一步扩大班级规模等等情况，为了遏止此种倾向，或许需要再强调一下我们一如既往的观点：教学要求应高于仅仅是教材"内容的传授"，为了达到这一目标，大体上要求学生-教师的比例应该接近于1。脱离这一水平均属不宜。

我们谨对下列诸位对本书手稿所提出的、收益良多的批评和建议表示感谢：Robert Auerbach, Douglas Braymer, Norman Hecht, John M. Hill, John L.Hubby, Patricia St. Lawrence, L. A. Snyder, Bernard Strauss, Paul B. Weisz, David J. West 和 Curtis A. Williams。Clare K. Woodward 先生对第五、六两章的合理化建设性意见亦使笔者获益非浅。还有许多在校大学生们全部或部分阅读本书后所提出的某些看法，启示笔者在安排全书内容方面亦作了新的考虑，对此一并谨致谢意。

Dow O. 伍德沃德
Val W. 伍德沃德

译 者 的 话

二十世纪下半叶，分子生物学开始作为一门独立的学科分支跃现于世。它的形成和发展是与物理学、化学、进化论与遗传学的进展分不开的。随着对大分子化合物的研究不断取得突破，特别是脱氧核糖核酸分子双螺旋结构模型的建立，人们能够从主要生命物质结构的分子层次上得以合理地解释基因复制的机理、信息传递的途径，阐明生物遗传变异的运动形态，从而使整个遗传学的研究跨入了以形态描述、逻辑推理为主转变为以物质结构与功能相统一为分析着眼点的新的发展阶段。

本书共分七章，翻译出版删去了第七章和索引。作者把生物化学、物理学、遗传学、进化论等各个学科分支在所谓生物信息流（信息传递）的总题目下，从时间（历史发展）与空间（信息流向）的角度比较系统地介绍了各类大分子化合物在生物遗传中所起的作用：它们具有怎样的分子结构，各类大分子化合物（DNA、RNA 和蛋白质）在结构与功能上的相互关系，它们在细胞中的活动方式等等。在本书第一章中，作者提出了教育改革的问题，主张高等学校遗传学的教学应着重于学生的思维训练和实验操作。并通过反对科学研究中的唯心主义哲学思想，阐述了对人类遗传学研究中某些问题的见解，特别指出过去将遗传学滥用于人类社会现象所造成的严重后果。其中即使某些观点有可商榷之处，但总的来说，它基本上反映出了作者的唯物主义认识论及其进步倾向。从第二章到第五章为全书的精髓所在。书中列举大量的具体实验和分析图表，集中表述了分子遗传学所特具的思维方法和试验手段，并以此为杠杆对遗传学的继承和发展，DNA 分子的结构与功能，信息的转录和表达，蛋白质的分子构型和生化特性，酶的作用机理及其调

控方式等，作了全面深入的剖析与评论。所引用的资料与已出版的同类书籍以及与有关的学科专著相比，其深度与广度，尤其是关于交换、干涉、蛋白质的结构、X射线衍射分析等方面的内容可以说是独树一帜的。因之这本《分子遗传学概论》虽名为概论，实质乃是一本较为完备、全面系统的教科书，而其所论述的具体范畴却又具有学术专著的某些特征。第六章专论进化，主要是推论遗传体制、即核酸、蛋白质相互关系的进化起源。虽然在叙述中推理多于实验，但新颖有趣，颇具特色，亦为现已出版的同类书籍中所不多见者。全书文笔流畅，图文并茂，极适合具有一般遗传学及生物化学基础知识的青年学生作教学用书或参考读物。

本书只能反映1976年以前分子遗传学发展的概貌。至于人工合成具有生物活性的基因（例如生长激素抑制因子的人工合成）、遗传工程的实施（如人体胰岛素基因在大肠杆菌中的表达）、左旋型DNA分子的发现等等，由于历史条件的限制，不能一一备载，只有通过查阅有关的专门文献，去求得补充了。

对于这样一部涉及专业繁多，浩瀚渊博的学科论著，囿于学术知识和翻译水平，我们在翻译过程中势必会遇到很多困难。承蒙多位学者、专家，特别是黄文几、刘祖洞、盛祖嘉、沈仁权、项维、高沛之、周叔彝、黄瑞章等先生在有关章节中给了我们极其有益的指导和帮助，才使我们得以完成这一工作，谨在此一并致谢。至于在译校方面可能发生的谬误和不足，欢迎广大读者批评指正。

一九八二年三月

目　录

译者的话
前言
第一章　生物学、科学与社会中的分子遗传学…… 1
第二章　遗传学简史……26
历史性介绍……26
基因和基因作用……43
DNA 及其在遗传中的作用……63
第三章　DNA……95
复制……95
突变…… 138
重组…… 157
第四章　RNA：转录与遗传信息的转译…… 180
转录…… 180
转译…… 194
遗传密码…… 216
蛋白质合成器…… 226
第五章　蛋白质：遗传信息转化为表现型的过程…… 246
氨基酸…… 247
蛋白质结构…… 257
酶的作用机理…… 304
酶活力的调节…… 318
酶合成的调节…… 330
第六章　进化…… 356
前生命进化…… 356

遗传器的进化…… 369
代谢途径的进化…… 384
基因组的扩增…… 386
蛋白质进化…… 398

第一章　生物学、科学与社会中的分子遗传学

生物教师在教学实践中与学生接触时，往往喜欢提出一些发人深省的问题来启发学生，一个司空见惯的题目是：区别生物与非生物的基本特征是什么？这一选题相当不错，只是在学生们思考了有关大象与自行车之间诸如繁殖过程、刺激反应、生长代谢、个体死亡等彼此间截然不同的几个特征之后，讨论往往就会到此为止。

能认真去分析大象与自行车有什么共同点的学生确实为数不多。至于拿出勇气去探索生命一词的含义，并根据生物由无生命物质构成这一事实，进而能深入思考这里面存在内在联系的学生更是寥若晨星。但是一旦能就生物与非生物的差别问题追本求源，立刻便会引伸出第二个问题：无生命物质究竟是如何相互作用才得以产生出细胞，并发展成为生命有机体的呢（图 3）？

把客观事物，例如包括从原子微粒直到银河系，进行分类整理对研究自然科学十分重要。生物体可分别归类为植物、动物及原生生物三类。这三大类群还可进一步往下细分，一般来说，植物界、动物界与原生生物界这三大门类内部，分类纲目比较明确。而对于那些三不搭界，地位含糊的对象，分类处理上颇费人猜测。例如，当发现病毒为其自身繁殖而离不开寄主细胞时，就会产生这样的问题：病毒是不是生命有机体，或只不过是复杂有机物的共聚体？人们在将病毒跟家兔、或甚至细菌经过一番比较之后发现，病毒与无生命系统的共同点，要较它与生物体的共同点为多。但是如将病毒与粘土颗粒进行对比，它那高度复杂的结构体制，却又显得更象生物体，而不象是非生物体。

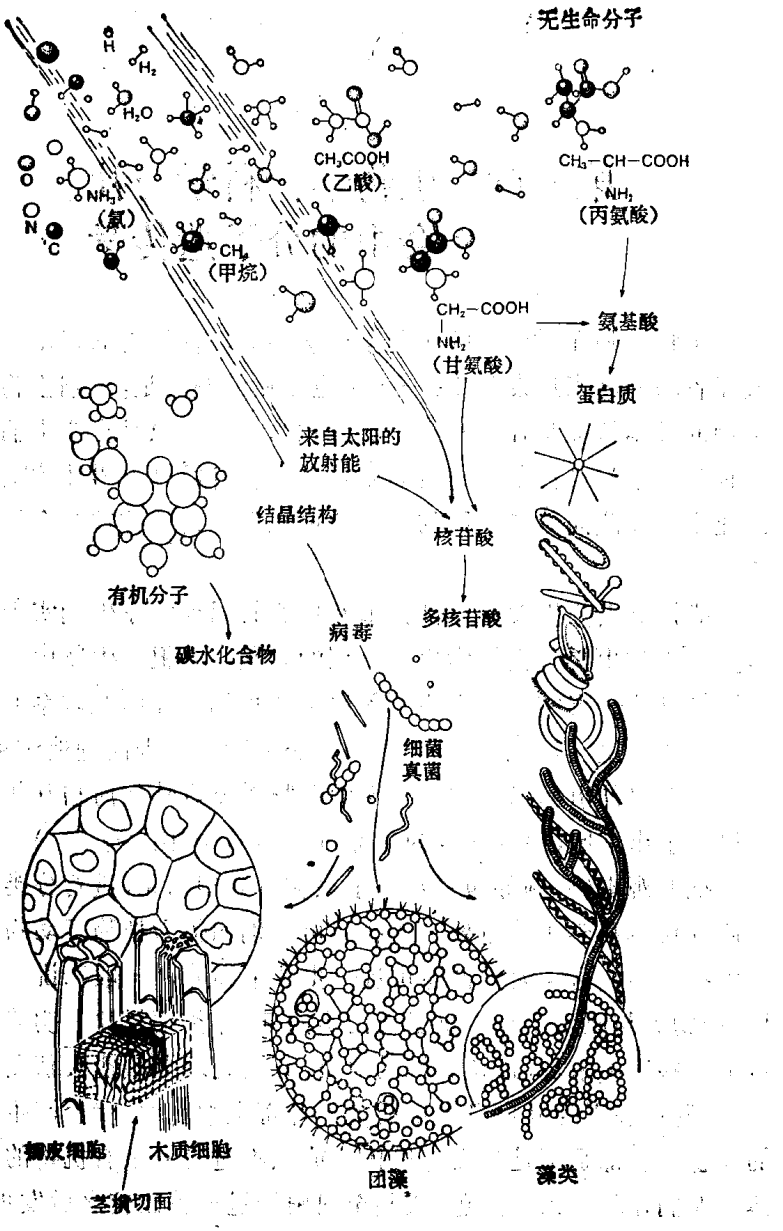


图 3 存在于非生物与生物之间的连续系列

分子不仅支配着生命系统的活动(功能),而且生命系统本身就全部是由分子组成的。在初始阶段,生命来自无生命分子的集聚和组合,而在现代生命系统中,生命活动也是通过无机分子转化为有机分子以及生命结构来维持的。综观整个过程,一条从最简单的原子,经过小分子化合物、大分子化合物、直到最复杂的生命系统的连续系列便呼之欲出了。

实际上在我们充分认识病毒的本质之前，争论病毒是生物体还是非生物体是徒劳无益的。最好还是考虑：病毒就是病毒。并就其形态本质作切实研究，而不必浪费时间硬性把它归入这一个或那一个分类范畴。那些坚持必须使生物系统中的每一要素符合于适当分类等级的人难免会发现，在全部要素一一吻合之前，继续分析相当困难。

病毒代表的只是我们在分类地位暧昧不清的研究对象中所碰到的一个困难例子。而确定生命或无生命的含义，或至少是对生命或无生命进行科学描述，则是一件较之更为复杂的难题〔参见克里克所著《分子与人》(Of Molecules and Men)一书〕*。第一种看法认为，"生命"与"活力"这两个用语没有什么意义。这一看法得到下述证据的支持：即从基本粒子、经历原子、分子、细胞、机体、群体等等，直到整个生物圈，现已表明基本概属于连续性系列。连续系列若断若续，并非由相互紧密相连的节段组成。生命与非生命之间有明显的界限鸿沟（图 3）。从连续系统的封闭性特征，可显示出在相邻各点之间存在着显而易见的相关性。例如蛇与蜴蜥、葡萄糖与半乳糖、DNA 与 RNA 等等。但是如果某二点在连续系列上彼此相距甚远，比方病毒与斑马、葡萄糖与细菌，乍看起来，可以说是毫不相干。因此只有在填空补缺的情况下，亦即透过查察完整无缺的连续系列，才有可能把在其中彼此隔开的对象挂起钩来。议论中得出的结论之一就是所谓"生命现象"并不是什么特殊现象，它只不过是"非生命"的理化现象的某种扩展而已。

与上述针锋相对的第二种看法是，生命本质上全然不同于非生命，两者极易区别，不容混淆。在历史发展过程中，此类意见曾表现为多种不同的形式。其中一派认为，生物体赋以的"生命力"(vital forces)是上帝以不同于"物质"(matter)的另一种形式创造出来的。生命力处于科学研究所能探索的范围之外，因之科学对于揭示生命的本质无能为力。

其中的另一个学派则采取唯物主义观点，认为人们对于全部

* 参考文献列于本章之末。

自然法则的了解，自然应该包括生命现象在内，可是有关支配生物体的法则却是绝对不可能通过物理规律以及化学规律来进行认识的。事实上最初把很多物理学家卷入分子遗传学这一研究领域中来的却正是这种对于开拓生物未知“规律”的异乎寻常的吸引力。在今天，生物科学研究中的活力论者已经是无足轻重了。也没有迹象表明，分子遗传学家中还有人会寻觅所谓生物学中的新法则，或谋求指在区分生命与非生命物质的新方法。虽然如此，但至今坚持那种把生物与非生物看成非此即彼、截然有别，并认为这一观点具有非凡价值的人，科学界还是大有人在。

对这一命题还有第三种看法。这就是综合上述几种论点，采其精华、兼收并蓄。主要根据是：一方面，显而易见，自然现象能够按其复杂程度，分类排队，予以等级化；另一方面也同样明确无误的是，大多数生物在质上有别于大多数非生物。虽然如此，生物本身却又总是由非生物物质聚集整合而成的。也就是说，在历史变迁过程中，生命是一个由多数无生命要素组合形成的复杂系统。

这一新看法除了令人注意到一类特殊的复杂等级系统（complexity）以外，并没有使问题变得简单。一个因素是等级迁变与任何动态过程一样，是难于捉摸和归档的。人们不可能将进化迁变分列为简显明白的分类等级，将客观事物进行分级往往会使它们看来象凝固不动。而实际上各个等级系统在历史迁变过程中总是处于变化、发展与进化的状态。

第二个因素是整体与组分的关系。例如氢原子的结构、功能、大小与红杉树有质的区别，但红杉树又部分地由氢原子组成。其结构、功能势必会受后者的制约。这意味着，氢原子与红杉树既同属于一个连续系列，但两者又明显具有表象直观上的种种区别。人们在研究氢原子与红杉树这二件客观事物时，既能够分隔开来，分别进行，又不妨作为一个共同系统内部的要素来综合分析。在考虑任何一个系统内部的无论哪一种要素的特性时，便产生了一个问题：处于等级变迁过程中各种要素的特性与它们游离于系统之外所表现的特性是不是完全相同？

人为的分类工作有时会造成某种分类体制貌似客观化或固定化的错觉。这是由于把混沌一团的客观事物切分为各类组分时(如把科学划分为化学、遗传学、分子遗传学等等场合)引起的。在某些情况下,分类的目的在于有利于对客观事物进行认识和交流;在另一些情况下,例如大学中分科设系则是为了加强行政效率。不过一旦确定某种分类体制,往往在开始时对于它能适用于哪许多方面也不是一眼就看得清楚的。

譬如,我们设身处地想象:设若我们是在乘儒勒·凡尔纳*的气球,以15米的高度作一次环球旅行。这时我们将能把植物毫无困难地跟动物区别开来,还能把动植物跟土壤、岩石与水流区分开来(图4)。从所处的这一特定位置,或许可以合理地把对地球的研究划分为植物学、动物学、水文学、地理学等等。进而设想,如果跳下汽球作近距离考察,不用说,立刻便会感到,人们对自然的概念必须大为开扩。这里至少还要加上真菌学、微生物学、化学等多种门类(图5)。不言而喻,随着人类观察自然的能力逐步改进,将会发现,对于自然现象还不得不愈来愈细地进行分门别类,直到最终建立起诸如象氢原子、氦原子、蔷薇类、鼠类、扁蓄类等细部组分为止。再冥想下去,如我们自身能减缩到氢原子大小,并从这一视野去剖析自然时,恐怕

图4 从15米低空俯视大地

从15米低空可以清楚无误地把大型动物与大型植物、流水及岩石等区分开来,视野中所看到的一切事物在一定的空间都存在着可能的相互关系(竞争、合作、互助等等)。(野生世界图片。)

* 儒勒·凡尔纳(Jules Verne),法国著名科学幻想小说作家。——译注

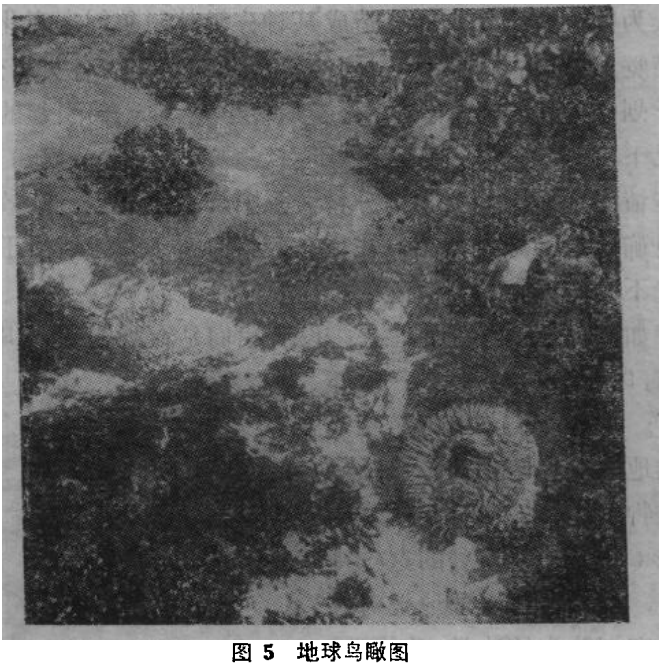


图 5　地球鸟瞰图

大地鸟瞰，表现出极大的景观差异。例如从赤道到两极，从荒漠到雨林，从海岸到山巅，自然生态千差万别。不同的图景启示着人们：对于物质与生命形态客观存在的复杂性应对之作出相应复杂的分析。(Nancy Hays, Monkmeyer。)

反而又分不出什么是植物、动物、细菌、岩石和水了(图 6)。不过，只要我们不采取这一极端立场，便总会要去寻找，并试图让别人也去寻找在整体的各个组分之间所存在的某种联系。一旦此种纽带变得鲜明耀目，人们将能从中看到不同事物之间，不同实体之间的相互作用，从而最终得以去研究物质的运动过程和各类系统的本质。

在寻找事物之间本质联系方面，分子遗传学很能说明问题，人们在分子遗传学渊源的问题上莫衷一是。不过有一点肯定无疑：它是早先就已存在的各个科学学科汇聚的产物。来自物理学、化学、生物化学、微生物学、遗传学及医学等各个不同领域的学者都为分子遗传学的创立贡献了自己的思想体系、技术手段和治学特色。他们的相互结合并非那种彼此一直无所关联的、以不同的门

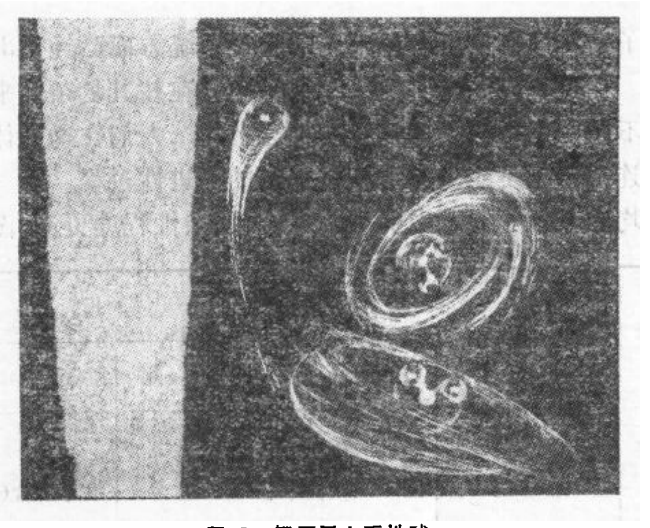


图 6　氢原子人看地球

设若把我们自己缩小到氢原子大小，则地球在人们眼中将是另一番光景：动植物差别将会消失，生物与无生物的界限亦不存在，人们所看到的将是一个广漠无边的空间，偶而亦有其他的原子闪烁其中。

类技术的简单集积。此种趋势的形成更多是属于这样一种社会现象：它吸引着人们以不同的角度和焦矩来审视存在于生物系统内部的信息流的问题。我们现在提及此点也是把它作为一种社会现象来考虑的，这是因为物理学家、化学家、遗传学家等等已经凝成了一个协约体制，并建立了互通情报的系统。在某种意义上说，这远远超出了他们原先所献身的学科系列和科学分支。实际上，它标志着一个新生学科分支的诞生，并从而导致了新旧学科系统和从事新旧学科科学家之间某种程度的分道扬镳。回顾一下历史上曾经把科学同行们简单划分为所谓正统派与非正统派的那段经历，再来体会现今的这种极向化和派系化现象，也许不无裨益。那样做的结果，使某些自命为正统派的学者对于科学技术发展中某些有趣的动向最终只能采取熟视无睹的态度。

分子遗传学的一部历史也是在解决下述诸方面问题上逐步改善视野的历史。即生物与非生物的实质差别是什么？在物质的最小单元与极其复杂的生态-生物系统之间究竟存在不存在连续系

列。分子遗传学在剖析各种不同的分类系统方面已经作出了微薄的贡献。在这里每一个分类系统都是建立在按其复杂性来说系属于各自不同的水平，亦即等级阶梯之上的。人们在这种体制上所理解的连续系列是能量与结构水平上的渐进性过程。后一级水平的系统均比其前一级水平的分类系统、复杂性程度更高(图 7)。

结构水平	表象项目			
	行为	进化	复制	信息流
生态系统	共生 竞争	竞争 互助	?	? 化学媒介(传递)
群体	社会等级	物种形成	?	?
生物体	恐惧、游嬉、学习	功能改变	有性，无性	基因型→表现型
细胞	运动	突变 选择	有丝分裂 减数分裂	DNA→RNA→ 蛋白质→表现型
细胞器	?	突变 选择	复制	DNA→RNA →蛋白质
大分子	?	突变	复制 (有模板，或无模板)	DNA→RNA →蛋白质
分子		化学变化	生物合成	化学活性
原子		H→He	?	反应特性

图 7 等级系统的划分

在自然结构的几个等级系统中，都有进化、行为这些项目，但支配这些表象的机理，在不同的等级系统可能不大相同。

在剖析全部的等级类型时所列表象项目完全相同。可以看到所有的等级类型在行为一栏上各具特色，并且有其本身的进化历史。其中复制一项对不同的等级水平都很重要。但是支配各项表象，如行为、进化、复制等等的机制在不同的等级水平之间则可能是，也必然是千差万别的。就进化这一项看，它按原子、分子、物种和种群的层次顺序展开。但如认为，用氢原子到氦原子的发展模

式也能说明从人类祖先到现代人的进化过程，甚至进而引伸于文化与经济领域，那就未免言不足据，失之妄诞了。

这并不意味着处于连续系统上的各个等级水平彼此间有着不可逾越的界限，或是质的差异使它们毫无共同之处。所要说明的只是对于每一级系统都需有各自合宜的分析与阐明其内在特性的方法。只有在通过直接方法，测定每一级水平的种种特性之后，人们才能充分地证实它们之间那种晦莫如深的类似性和相关性。此外，客观存在的进化现象当然也能够说明这种关系。

达尔文虽然不是第一个强调自然事物本质上是动力学的、变化的观点的学者，但他无疑是就"运动"含义，也就是现今所理解的物种进化正式作出解释的首批作者之一。达尔文提出，所有生物都由它们共同的祖先把彼此相连在一起。也就是说，当向上追溯生物的发展历史时，总会发现不同的生物类型最终会在某一点上汇合起来。至于说生物的历史源流是会集于一点，还是有几个起始点的问题，达尔文并不穷于对付。他的确揣想过，从一个起始点，还是从几个起始点都有可能发展成为为数众多的新型物种。在有关物种进化的正式说明中，达尔文主要将之归纳为自然选择与自然变异两方面的内容。

自然选择这一述语常常用来阐明那种包罗一切的迫使整个物种个体在成熟前死灭的自然力量。而自然变异则是指人们所观察到的在物种内部所存在的不同个体之间的差异，以及某些个体要比另一些个体有更多的生存机会，或者更容易被淘汰。如把达尔文学说高度概括起来，就是：一个物种内部的全部个体在遗传上是不尽一致的，亦即个体之间广泛存在着本质上的变异。达尔文认为这种情况必然会导致一些类型比另一些类型更适于生存。那怕是新生个体在存活潜力上的差异微不足道，但在时间的流程中自然选择依然能觉察此种差异，并使某些个体归于绝灭，而让另一些个体扩大增衍，大量繁殖后代。

达尔文本人并不了解现代遗传学理论，但他对现代遗传学中的一些概念，如突变、遗传饰变等却表现出惊人的先知先觉。达尔

文曾使用微变异这一用语记述了今人周知的基因突变现象（它可以解释一个物种内部的个体间变异，例如眼色、体高、血型等等）。在生物信息流概念方面，也作出过令人叹为观止的预言。对于生物发展史，达尔文同样有着敏锐过人的洞察力，并注意到了任何物种都保存有其本身衍变，以及与其他种属间相互关系的历史痕迹。微变异或突变可以认为是物种衍变历史记录中的海滩一贝。事实上研究物种进化也就是在研究生物信息：研究生物信息是如何肖规曹随式地往下传递的？历史印记又是怎样通过突变发生改变的？研究自然选择又是如何地发挥作用，使一些突变得以保存，而使另一些突变湮没无存。

生物进化的概念无论如何要比图 7 表示的范围广泛复杂得多。从图 7 可见，进化一栏在各个等级水平上全有表现。在生物系统外部或是在生物系统内部，无机分子、有机分子、大分子都在运动发展。在这里达尔文的阐明物种进化的理论并不能解释氦原子是如何从氢原子"进化"的，也不能说明氨基酸、单糖等又是如何从水分子、氨气、甲烷加上能源形成的。可见需要另外一种完全不同的理论解释，是势所必然。

除了进化、行为、复制等方面以外，表象项目中还有信息流的问题。在每一级结构水平上，人们看到信息会从一个地方或一个实体传递给另一个地方或另一个实体。一个原子中包含有的电子数目便为该原子提供了极其重要的信息，某一群体内部的个体类型及其数目同样也是信息。它展示着该群体的发展进程的过去和未来。达尔文的进化论只是试图努力在物种的水平上，而不是在个体、细胞或大分子化合物水平上来记述此种信息流。然而，只要就个体、细胞及大分子的等级水平作一番仔细考查，就不难理解，这里确实存在着某种共同的机理，推动着信息流在上述各类等级水平内部的相互传递。

本书即着眼于大分子化合物、细胞、机体水平上的信息流问题。并在讨论信息流时将之解离为一些较为简单的过程，包括信息的贮存、复制、转录、转译、生物信息的调控，以及从信息转化为

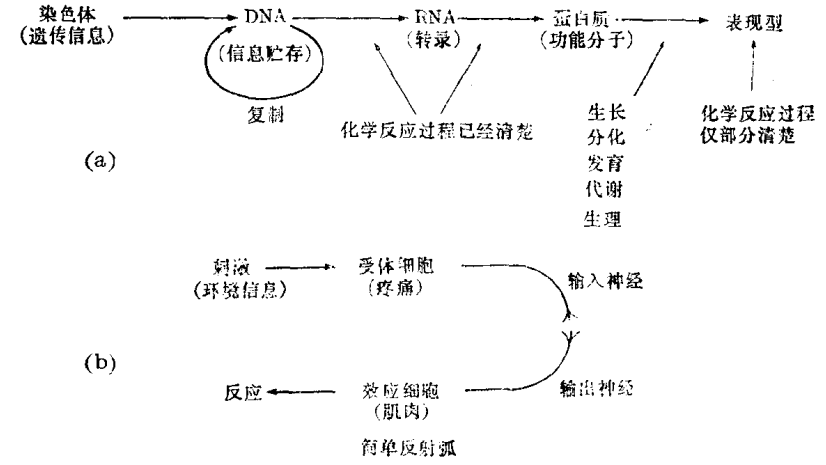


图 8　生物系统中的信息流

信息流在生物系统内部的传递涉及到一系列机制：(a)通过染色体上的基因把遗传信息一代传给一代。而且在同一世代的细胞或机体内部，遗传信息亦可从基因传给 RNA，再到蛋白质。(b)环境信息同样能在生物系统内部进行传递，例如环境信息可由受体细胞察收，经过神经传递，并通过肌肉，内分泌腺等转化为动作。

表现型〔图 8，机体的表现型涉及外观、物理、化学特性及(或是)行为等等〕。目前有关信息流的生物学机理已经大白于世。至少在上面图示的层次上确是如此。本书的目的是想把图 8 信息流所示的概念图解进一步放大，并从细节上详加论述。

分子遗传学在于专门研究生物系统内部信息流的表现。它论述的范围及重点稍有偏狭，因之不可避免地会遗漏信息流讨论中许多重要而有兴味的问题。但另一方面，本书也搜集了不少最新的，具有重大意义的尖端性内容。我们在这里把主要的着眼点放在所谓信息大分子的一类大分子化合物的结构与功能的关系上。具体指的就是 DNA、RNA 和蛋白质(图 9)。从严格的遗传学观点来看这三类大分子化合物正是遗传信息流的最基本的负荷者。

DNA（脱氧核糖核酸）具有贮存、复制和转录遗传信息的能力。DNA 的结构与它所担负的复制和转录的功能是完全协调一致的。此种情况意味着，可能是在功能的基础上使 DNA 的结构

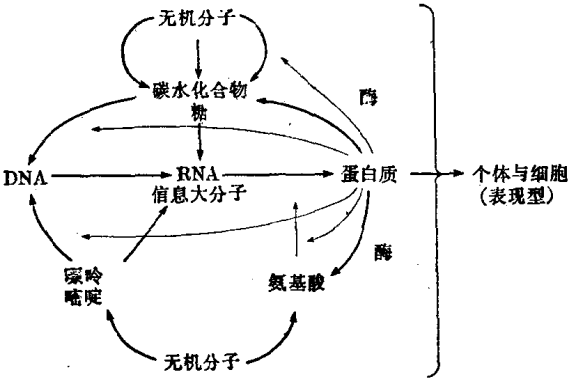


图 9　信息大分子

上面有关信息流程的这张蓝图不仅描述了各信息大分子间的关系，而且也勾划出了本书内容的概貌。总之，本书的主题就是阐述信息流从 DNA 转递到 RNA 再到蛋白质的机理。每个箭头均代表一系列化学反应。

达到合理化。结果从 DNA 的结构特点正好解释 DNA 所特具的功能。一般来说，RNA 亦复如此。这是理所当然的。这里转录是指遗传信息从 DNA 转递到 RNA 的过程。这一过程从 DNA 与 RNA 的分子构型同样能予以说明。RNA 分子合成时总是以 DNA 作模板。因之 DNA 模板的结构与成分便决定了 RNA 转录本的结构和成分。信息从 DNA“磁带”转录到 RNA“蓝图”上的过程不妨比喻为音乐家灌制唱片，当然各自的机制是相距天涯的。

蛋白质分子也包含有遗传信息，但蛋白质不能象 RNA 那样从 DNA 直接获得信息，而是通过 RNA 分子把其上的信息翻译成蛋白质信息贮存起来。尽管现在了解，信息从 DNA，亦即从基因先转录到 RNA 分子上，再由 RNA 分子翻译成蛋白质结构，最后在功能上得到表达这一确切无误的途径，但是对于蛋白质结构和功能之间翔实可靠的关系我们迄今还是不得而知。与在核酸研究中我们就其结构与功能所取得的知识相比，简直有天渊之别。也许正是由于对蛋白质结构与功能关系上这种未臻完备的认识，使它们本来就存在于许多分子遗传学家心中的魅力，更增添了无法抗拒的诱人色采。

在信息贮存、复制、转录及翻译方面有不少问题尚待进一步阐明。但即便如此，总的轮廓目前已经清楚，并有可能就上述过程提出大量细节说明。但是，如果我们沿着图例中生物信息从 DNA 到蛋白质再往下推敲，就会发现，要完满地解释许多细胞及机体的表现型还有很大一段距离。换句话说，对于包含在 DNA 分子中的信息，最终确认会转递到蛋白质分子上，可是从整套的"信息蛋白质"又是如何发育成表现型的问题目前依然一无所知。所以，我们在本书中不拟过多地在信息怎样从蛋白质转为完整表现型方面花费笔墨，而是把中心主题集中在信息从 DNA 向蛋白质转递的问题上。

与任何知识一样，在生物系统内部的信息方面获得的最新知识也正在被应用于"控制人类环境"。今天科学知识已为切实控制遗传信息流提供了一定程度的可能性。在某些情况下借助于对原始"录音带"(DNA)作特定的更动，以便使"信使"发生改变，甚至还能够在试管中人工制造出新的录音带，随后将其整合到活细胞的染色体上去(有时亦称作遗传外科手术)。人为干预和改变信息流的能力目前已经引起不少公共出版物和专门刊物的注意，并就其重要意义作了大量评论。这方面包括有越来越多的科学家正在慎重衡量，在制定出某种道德、政治上的法规之前推迟此项工作所可能造成的功过得失。可以断言，围绕着应用信息流的科技问题要想作出一项道德和法律章程尚属遥遥无期。目前的危险可能就在于有人企图先斩后奏，在原则性的意见公布之前，开创先例，造成木已成舟的局面。

一本分子遗传学教科书并不是宣传应用科学知识的适宜场所。许多科学家甚至主张不论是教科书，还是科学杂志和公共出版物都不应鼓吹这个问题。对多数人来说，这里或多或少有个约定俗成的文法。这就是说，科学家的用武之地应该是实验室，因之在有关应用这一科技知识方面的讨论，以及就社会对科学活动管制问题上的种种争议全都是"不务正业"。此种观点与自由探讨、公开辩论以及客观主义等概念是近乎对立的。

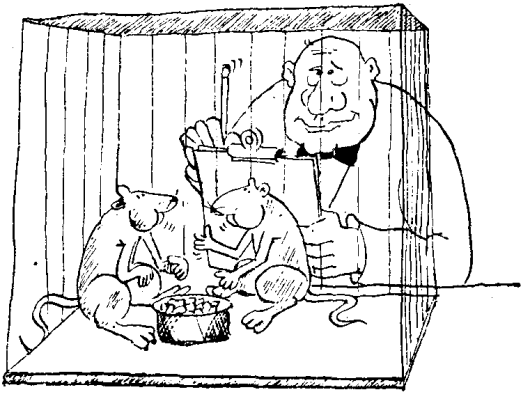
某些科学家不愿意公开讨论这个问题，部分原因是由于社会上有相当多的人长期以来对自然科学持有一些不正确的看法。其中一种误解是把科学当成是绝对纯粹的自主的世外桃源。亦即自然科学与人们普遍关心的另一种意识形态如伦理、道理、政治等等完全无关。在科学家中和不从事科学工作的社会公众中，这两方面都有不少人认为，现代科学已经成为一支与自然"对话"的专用热线。从自然界受到的信号无疑是客观的。因之科学家的社会责任仅仅就是要在那些有声望的杂志上发表每一项与自然"对话"的准确无误的记录。

还有一种偏见，即认为科学具有不偏不倚的"中性"属性。此种看法的来龙去脉至今无从稽考。它的产生或许与这种偏信不无关系，即所谓在"纯理论"科学和"应用"科学之间有着不可逾越的质的差别。一般说的中性科学指的就是"纯理论"科学。奉行此种信念的理由是"纯理论"科学探求的只是基于好奇心而提出的那种毫无偏见的"纯粹知识"。目的只在于探索自然奥秘，而不是谋求为我所用。这也就是为什么农业和医学院校与作为基础理论科学的生物系、生物化学系等系科保持一定距离的实质所在。有些从事于"纯理论"学科的科学家常常傲慢地把应用科学问题贬低为世俗的雕虫小技。而另一方面，应用科学家们也反唇相讥，斥责专搞"纯理论"的科学家是一群脱离实际、不食人间烟火的云端神仙。

任何理论、资料、事实都来自于一定的目的和活动。反转来它又能据以提出更深一步的目标和预测。例如医药和农业就是部分地在理论指导下形成的。而理论本身则又来自于实践。可是客观现实却使"纯理论"科学和"应用科学"相互隔离到如此荒谬的程度，以至达到人为编造出所谓中性概念的神话。在美国今天几乎所有的科学研究都由州政府、联邦政府乃至私人基金会提供财务资助。作为条件，一切研究计划都需要先提出证据，说明它在医学、农业、军事等各方面所可能预期的实用价值。科学与技术知识水平往往是构成一个国家政府、工业和军事力量的基本源泉。这意味着，在科学与社会、知识与力量以及科学思想体系与政治意识

形态之间，确实存在着不可分割的现实联系。

很明显，对 DNA、RNA 和蛋白质开展的研究，乍看起来是与指导人们和集团每天活动的政治因素和社会力量两相无涉的。但仔细查看一下就会了解，有关大分子信息与生物信息知识的运用，目前还存在有各种争论。在今天，“遗传工程”（从优生学到“基因外科”、到建立基因无性繁殖系）、文化工程〔从生物反馈到改变习性的反感疗法(aversion therapy)〕实际上已经成为当代社会中引人注目的潮流(图 10)。因此在我们看来，使学生能够充分认识科



行为限制：“要看我怎样操纵这家伙吗？
只要吃一匙食物他就会机械地在纸上划个记号。”

图 10　人在生物反馈作用下的行为限制

学知识的重大意义，了解科学知识在工业、政治与军事方面的作用，以及科学知识本身决不是不偏不倚的中性或是游离于社会政治的一个独立的范畴，这些都是全体科学家们不可推卸的责任。

对于那些一心想成为科学家的学生来说，要是有机会了解社会政治与自然科学相互关系的后果，他们很可能便不愿干某一类工作。但如果他们能认识到人们必须为赢得生存而进行研究，而且不言而喻还需要从某一方面接受经费才好开展研究，上述情况便不大会发生了。同样，对于并不冀求将来做科学家的学生来说，如果使之在掌握自然科学的基础上进一步学习科技知识的应用，他们也会乐意选择作为一个普通人去参加建设设计的组织工作。

自然科学在私人生活与公众生活中都有其十分重要的地位。例如在私人生活中对于有些自认为可信的关系和证据等等，科学能够以"新的思维方法"，提供重新恰切研究的余地。此外在保证个人健康与福利方面，科学无疑也会给予有益的指导。当然并非所有的科学家都同意"新的思维方法"是尽人可教的。多数人似乎感到，创造性只是属于那些赋有创新能力的人。而现实生活中的大多数人应该满足于平庸的生活与轻松的工作。不论此种看法是对是错，要是人们所指的基本前提是说，教育应该能有助于提高一切人的生活技能，上面的这番主张便驴头不对马嘴了。大家知道，人们是在一个技术社会中生活，因之绝大多数人都完全同意，在今天如果要得到并保持那些为个人生活所必须的工作手段，他就必须非懂得知识技术不可。基于这一缘由，讲授科学知识的教师是有能力在这方面有所作为的。至少在我们看来，有可能做到使学生能够熟悉自然科学的作用，并如何在实践中应用科学知识方面有更敏锐的洞察力。但是如果对科学知识的讲授，其重点只限于科学资料与实验方法，便不易达到这一目标。

除了所谓科学自主性与科学中性论等神话之外，还有一些妨碍如实地教授自然科学的因素。其中有些与没有给学生及时讲授自然科学中各种认识论的基本概念与特色有关。——这个问题本来是应该讲授的。有不少人在这一点上再次表示不同意我们的看法。他们既不赞成我们的分析，也不赞同我们所一直坚持的主张：科学讲授的范围最好比指定的科学知识要更广泛一些。

现代科学在认识论上有三大特色。这就是还原论、逻辑实证论与客观主义。有关这些方法论的问题并不是自然科学固有的属性，也不是科学研究本质性的内涵。目前对于它们真正确切的含义，以及它们对科学与社会的影响依然缺乏充分一致的认识。可是正因为它们对自然科学与自然科学家确实在起着作用，所以对于它们的含义和作用作一番认真的了解十分重要。

生物学中的还原论表现为两大形式。第一类形式往往视为在所难免。这就是为了分离与研究一个复杂体系（例如一个细胞或

一只小鼠)中的某些较为简单的要素,如线粒体、肾上腺,而将该复杂体系进行分解破碎时所随同而来的信息遗失(消减)。诚然,一个复杂的系统要比较为简单的系统在研究上更为困难。但是,如果为了研究的方便而把一个复杂的系统予以简化分解,则同时又不免带来信息遗失,甚至从复杂系统所分离出的一个或几个要素特性被测知以后,还不一定就真正说明这些要素在复杂系统之中表现的真实特性。每一种要素都在通过与同一系统中的其他要素相互制约而改变着其他要素的作用,与此同时,其本身亦不断地被其他的要素所修饰。在一定的场合,所遗失的信息虽然有可能会借助于系统的重建重新获得,但是在一般情况下,只有对系统本身进行分析才会掌握该系统的真正特性。那种把各种要素的个别特性累加起来似乎即成为系统本身的做法是不可能反映一个复杂系统的本来面目的。

还原论的第二种形式可以说是渊源于第一种情况。但是更具有主观武断性的色彩。这一类还原论本质上认为,处于等级阶梯上较高阶段的高次现象可以通过低次现象的运动形态来取得认识。例如有人断言,生命仅仅不过是各种物理、化学现象的总汇。而对于物理学家和化学家来说,这些现象其实是众所周知的。这是分子遗传学中运用这一逻辑系统的一个明例。

的确,在微生物中有不少结构-功能-行为性状可以用鲜明确切的机理模型来作解释。通常在分子遗传学中的机理模型是这样的,即能表明细胞或机体的每一种特性都是基因作用的直接结果,亦即基因决定性状,称为遗传决定论。细菌中有很多性状能归之于特定基因的作用,这在今天是无庸置疑的事实。此外还有明显的证据,表明人类的性状也同样好用特定基因的作用来说明。但并不是说所有的细菌性状都是一对一的基因作用产物。而且即便在细菌中一切性状是遗传决定的,也不能就依此类推(以准确的科学性),认为人的特性表现一定也是如此。换句话说,认识细菌的遗传系统并不等同于也认识了人的遗传系统。当然实际情况或许可能这样:人的遗传系统与细菌的遗传系统二者本质上不相上下。

但是要确证此点，还需要对人类遗传系统作深一步试验研究。借助于此类研究，我们便可据以与外推分析法的结果作对比。而目前关于人类遗传系统的信息表现多数都是从细菌遗传系统的信息表达沿用外推法来推论得知的。事实上在人类各类性状上运用还原论逻辑系统的例子举不胜举。值得指出的是，这些都是在对此类性状远未详尽研究的情况下借助外推法而作出的判断（图 11）。

由于还原论逻辑系统强加给人这样的观念——人类行为中有许多都是遗传决定的，结果使我们社会中的不少机构亦以此作为工作依据，克尽职守。教育实践则是在两种观点之间摇摆徘徊。一种观点是“智力取决于基因”；另一种观点是“智力是实践经验的结晶”。在第一种场合（智力取决于基因），似乎由遗传规定的性状是固定不变的，不可能通过环境干预使之改变。在第二种场合（智力来自于经验），教育系统则力求为学生提供得以提高智力的经验，但是是否一定导致所希望的结果，对此并无十分把握。就还原论的客观效果来说，不仅教育系统成了这一逻辑系统的牺牲品，而且尽人皆知的是，公道概念与司法系统几乎也是因人而殊的。不管实际情况究竟如何，在对待所谓的“野蛮人”时，与对“文明人”就是二副面孔。

在序言部分曾简略提到，逻辑实证论的概念对于人们进行科学判断很有影响。这一概念的主要精神是要求科学资料必须与实际情况相统一。按照逻辑实证论者的逻辑推理，一切事物必须一分为二：一方面是客观存在的资料，另一方面是主观形成的观念。自然世界由客观事物构成。在对客观事物的研究之后会积累资料，但研究过程与主观思维又密不可分。因之通过思维与感觉形成的观念也是主观性的。就这一方面来说，逻辑实证论与客观主义在整体上有其相互关联性。

根据实际获得的资料进行思维、分析、归纳是自然科学研究中一大重要特色。获得资料的方法多种多样，科学实验是其中之一。要教育在实验室中工作的科学家在实验过程中应力求避免偏颇与差错。在资料收集过程中必须保证数据客观性是科学实验的首要

图 11　用还原论逻辑系统看人类行为

由于草原犬鼠的分布是有区域性的，便因而说，民族主义亦有其生物学的本源，这一结论是否正确呢？如果对民族主义的整个社会基础缺乏充分了解，似乎就没有根据作出这样的结论。当还原论的逻辑系统试图规避旨在充分认识某一复杂现象的全过程，而把那些较为简单的系统的特征说成是这一复杂现象的特征时，人们或许会产生这种印象：作为我们大家所需要真正认识的人，与人所熟知的猕猴、细菌、变形虫等等岂不过在伯仲之间了。当然还原论的逻辑系统会有助于从复杂事物中汲取次一级组分供仔细研究。这一点在科学研究中仍不失为一种行之有效的方法。（上图由 L. Rue 与 T.Fujihara 馈赠，下图由 L. Rue 与 M. Forsyth, Monkmeyer 提供。）

前提。在这一问题上，科学组织内部的意见是绝对一致、毫不含糊的。所以对大多数科学家来说，无论在其实验结果公布之前，还是以后，依然还有一个追本穷源、继续审核的过程。总之必须尽一切可能杜绝实验差错。

然而，进行实验、收集数据只不过是整个科研过程的一个部

分。另一重要阶段包括在科学的基础上形成概念和创立学说。作为理论和学说，例如进化论(这个问题是很早就一直在争论的)其本身已经包含有大量的客观事实以及从客观实验与自然观察中所得出的种种结果。可是逻辑实证论者在对之进行概述评价时，由于他们使用客观主义作衡量尺度，于是在他们眼里，这些学说便都毫无例外地沦为主观性的思维形态。因之不可避免地陷入观念范畴。事实上一个理论在未曾经过客观资料的"证实"与"否决"之前，很少能被指称为是正确的理论，或是错误的理论。只有当它与大量事实不能再相适应，和对之无法作出合理解释时，这样的理论才算是真正过时。在通常情况下，这种过时的理论必将会被新的理论所代替。但是按照实证论的"逻辑"，即便是一个杰出的学说也仍然会被视为主观性的东西。在现实生活中，由于认识上不全面，在此种压力下，分析事物应该要尽量做到"客观决定"(这是因为客观资料太贫乏了)，这是不言而喻的。但另一方面，"不完备"又是相对的，其范围可以从 0 到接近于 1。人们不应对此过分强求。总之，逻辑实证论者在划分主观性与客观性的界限方面所花的精力，比之如何学习去处理各方面的问题，以加强客观资料的收集与主观分析的能力，恐怕还要更多。

任何一个科学家，他们在收集资料的工作阶段，从主观上说，务需竭尽所能，力争客观。但不同的科学研究，其收集资料的过程很不一样。为了说明这一点，我们来考虑一下各门科学所固有的数量性质与质量性质。例如，把分子遗传学与物理学、化学作一番对比，就可知道，前者很大程度上属于质量性质的自然科学，后二者则具有更多的数量性质。质量性质的"资料"，例如高和矮，罹病与正常等等，客观表现上泾渭分明，不会两相混淆。很明显这是由"感觉资料"本身特具的截然可分的性质所规定的。质量资料同时还能掩盖数量性质的表现。(在不少早期的遗传学教科书中常说基因决定着人类眼色。当一个人存在正常基因 B 时，为棕色眼，如具有该基因的突变基因 b 时，便为蓝色眼。这里总的倾向是把人群分为棕眼与蓝眼两大类。其实在人群内眼色变异，范围甚广。影

响人眼色泽的基因亦相当可观。但由于蓝眼与棕眼相当明显的区别，结果便掩盖了眼色上的许多细微差异。）事实上，被掩盖的数量现象照样具有各自的全部表象，只是在人们主观上常常视而不见罢了。

在科学研究中，科学家选做哪些科学实验，参加哪些科研项目，以及在研究过程中是追索其质量性质呢？还是分析其数量性质？这方面的主观色采可能更为浓厚。科学研究与衣着、言词以及生活消费一样往往也有赶时髦的倾象，科学家也象普通人一样地想赶时髦。但即使这样，科学中的时髦也与社会中其他领域的时髦不尽相同。重要的是，应该在他自己的那一学科部门中，以及在支持科学研究的主管机构中去识别那些真正懂行的人。另一点重要的，则是应该献身于某个"时兴的学科"，最好是创建一个新的学科分支。这是因为，在科学家中以及在各种学科之间还存在着如图 12 所示的那种社会高低的等级。

科学研究中存在有某种程度的主观性，并不等于说科学便因之失之谬误。主观性本身不是错误，更不是不诚实或偏颇的同义语。一切科学研究从发现客观规律到创立理论，或是提出一种研究假说，全部应归功于科学家中枢神经系统的思维活动。科学资料与其他资料一样也是感觉资料。通过对感觉资料的持续实验，将会在一定范围内取得愈来愈高的内在一致性。例如电子学或分子遗传学等学科就是如此形成与发展的。内在一致性与客观性仍不能等量齐观。一般只要排除系统内部的噪音干扰便不难达到这种内在一致性。而在对什么是噪音，什么不是噪音，进行决择取舍时常常也带有主观成分。这就是说，形成内在一致的过程本身就是主观决定的。

科学是人类的事业，因之也象其他的人类活动一样，必然会受到占主导地位的社会观念形态的支配。西方科学采取放任主义的态度很可说明问题。科学研究是在一定的社会制度的庇护与资助下进行的。特定的社会制度将在很大程度上规定着科学竞争的法则与科学研究的条件。一般来说，竞争是极其尖锐的。不少科

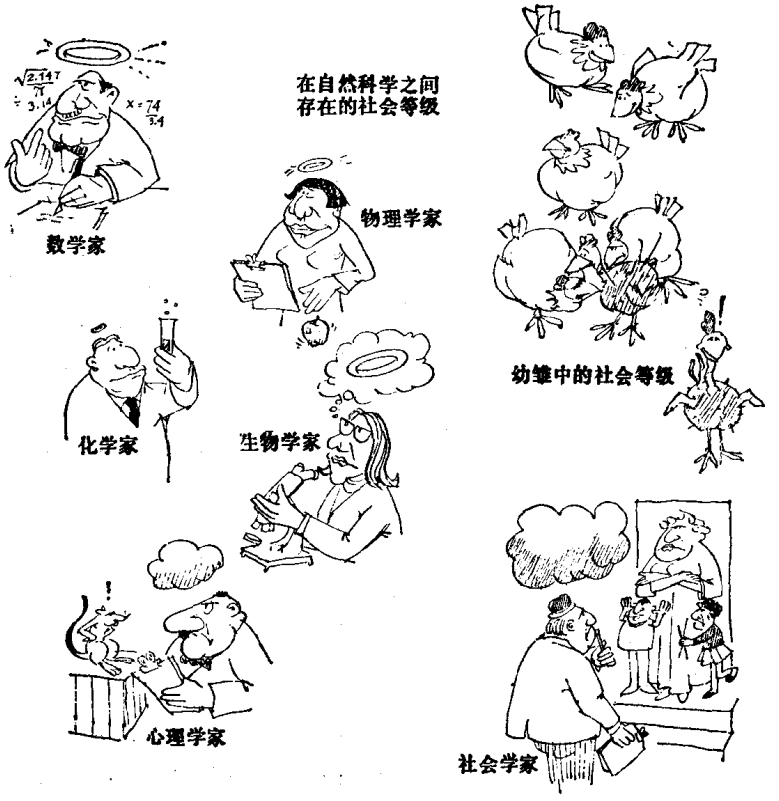


图 12　还原论者对科学家与科学学科中所存在的社会等级的解释

物理学、逻辑学和数学常被人们视为科学等级的上层学科。下面一级是化学，再下面是生物学，依此类推。但是按每门学科所研究的系统的复杂性(与重要性?)的"等级顺序"来说，这一传统的等级观念，在相互关系上恰好是本末倒置了。

学家都在不遗余力地寻求奖赏和赞誉。更多的人则致力于能得到高层人士和政府机关的赞助。其实，在这一点上科学家和科学机关与一般人和普通机构相比，亦无特别出格之处。可是最近以来，自然科学与自然科学家的公众形象却一直在为此遭到舆论的非议和诘难(阿尔伯特·爱因斯坦是一个典型，Jekyll 博士是另一种典型)(图 13)，这便不大公平了。

科学研究中的时髦花样本身就是层出不穷的。这里包括使用新字眼，创造新名词，买弄新风格以及科学家个人的固执偏见和标

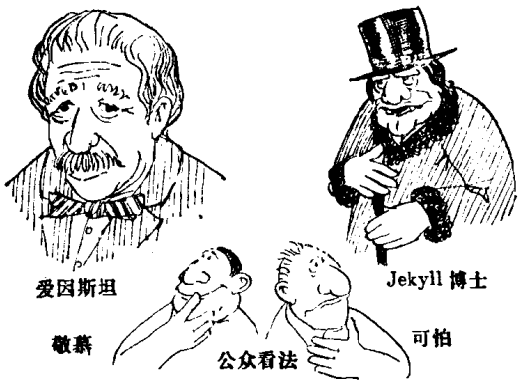


图 13　一般人眼中的自然科学家……

新立异等等。而后者则往往会在科学事业中带来新的突破。例如，孟德尔在说明其豌豆遗传模型的论文中，提及有关的遗传单位时曾使用“因子”一词。多年之后孟德尔的后继人为了纪念孟德尔的功绩，把因子冠以孟德尔的名字，称“孟德尔因子”。但是还在孟德尔的试验再发现并得到广泛证实后不久，W. 约翰逊就已开始用“基因”来代替孟德尔因子这一用语了。基因一词使一大批生物界后起之秀兴趣倍增，他们都是从那些缺乏想像力的平庸僵化的同行中脱颖而出的新人。同样的情况在 50 多年之后又重演了一次。这就是分子生物学与分子遗传学的新名词对另一批生物学家的有益影响。早期的遗传学家常被世人喻为“基因猎手”(gene chasers)。无独有偶，分子生物学家被他们的化学家同事讥喻为“无执照的生化实干家”则是 1950 年后十年间的事情。

最近几年分子遗传学风靡一时，名闻遐迩。但即便如此，关于分子遗传学用语的来源及其含义，迄今意见仍有分歧。不少卓然杰出的生物学家宣称，无疑正是这一用语特具的魅力才使科学家们聚拢到一起，形成了一个新的“时尚的”科学研究组合。所有这一切都应归功于人们创造了一个新的科学用词——分子遗传学。大量证据〔参见 E. L. Hess 的专著：《分子生物学的起源》(Origins of Molecular Biology)〕表明，在从事分子遗传学研究的科学家

们中间都强烈感到，生物科学研究中的一个完全不同于往昔的全新的启蒙时代即将到来。当然并不是人人都会为之感到欢欣鼓舞的。无论如何，至少传统的生化学家还没有意识到迫切需要一个专门研究生命系统中的分子及其化学反应过程——而这本来就是传统生化学家们的世袭领地——的新学科分支。另一方面，传统的遗传学家在认识上又同样“因循守旧”，把有关基因的化学结构与功能问题只放在他们研究日程上比较次要的地位。

上述的传统观念一般不大注意这样的事实：除了遗传学家与生化学家之外，其他学科的科学家们也提出某些与之有关的问题，并急于寻找它的答案。物理学家、化学家和微生物学家们都在为遗传物质(基因)的结构与功能问题、信息流从基因向表现型传递的问题等等大为操心。大家群心群力，不仅为分子遗传学摆出了许多被认为是饶有兴味的问题，而且更进一步提供了新手段、新思路，从而使这一学科终于取得人们的偏爱与重视。这一切在一定程度上开始形成了某种学科的联系。并据以大大推进了生物信息

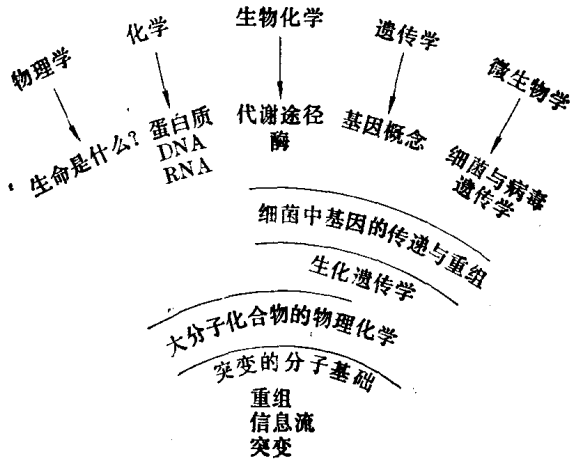


图 14　分子遗传学的来龙去脉

本谱系至少说明三大要点：(1)分子遗传学的起源问题是极其复杂的，很难对之究本穷源；(2)分子遗传学从许多科学分支接受丰富的遗产。因之(3)任何追索分子遗传学由来的图谋只会暴露目前各学科之间界限的人为性质。从本谱系表可能只会看到有关分子遗传学起源的很小一部分事件及其相互作用的情况，而且，不同的生物学家对分子遗传学的谱系问题必定会有各自不同的版本。

流问题的研究。由此看来，分子遗传学也谈不上是什么崭新的学科分支，它不过是多种学科在偏爱的基础上的结合。也就是说，在此之前科学研究各个领域中原本各行其事的诸种学科现在在分子遗传学这一总的专题之下被有机地综合在一起了(图 14)。

参 考 文 献

一般

Davis, B. D., R. Dulbecco, H. N. Eisen, H. S. Ginsberg, and W. B. Wood: *Microbiology*, 2d ed., Harper & Row, New York, 1973.

DuPraw, E. J.: *Cell and Molecular Biology*, Academic, New York, 1968.

Goodenough, U., and R. Levine: *Genetics*, Holt, New York, 1974.

Lehninger, A. L.: *Biochemistry*, 2d ed., Worth, New York, 1975.

Luria, S. E.: *36 Lectures in Biology*, MIT Press, Cambridge, Mass., 1965.

Mahler, H. R., and E. H. Cordes: *Biological Chemistry*, 2d ed., Harper & Row, New York, 1971.

Stent, G. S : *Molecular Genetics*, Freeman, San Francisco, 1971.

Watson, J. D.: *Molecular Biology of the Gene*, 3d ed., W. Benjamin, New York, 1976.

序言和章节

Ansubel, F., J. Beckwith, and K. Janssen: The Politics of Genetic Engineering: Who Decides Who Is Defective? *Psychol. Today*, 8(1) (1974).

Cairns, J., G. S. Stent, and J. D. Watson, (eds.) *Phage and the Origins of Molecular Biology*, Cold Spring Harbor Laboratory of Quantitative Biology, Cold Spring Harbor, N.Y., 1966.

Chargaff, E.: *Essays on Nucleic Acids*, American Elsevier, New York, 1963.

Crick, Francis: *Of Molecules and Men*, University of Washington Press, Seattle, 1966.

Freire, P.: *Pedagogy of the Oppressed*, Seabury, New York, 1971.

Hess, E. L.: Origins of Molecular Biology, *Science*, **168**:664 (1970).

Morris, D.: *The Naked Ape*, McGraw-Hill, New York, 1968.

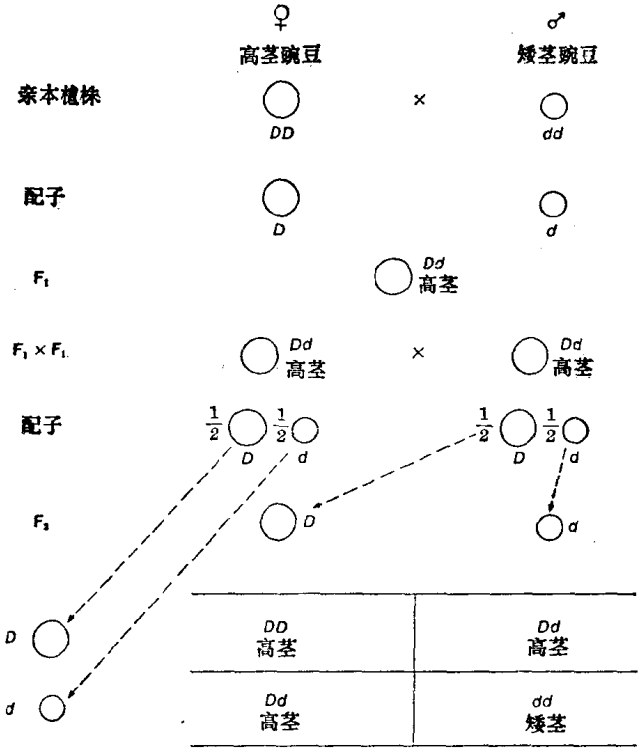
第二章　遗传学简史

历史性介绍

化学、遗传学和进化论这三门重要学科大大地推动了分子遗传学的发展。在十九世纪，化学和进化学说已成为完全成熟的学科分支，然而，遗传学则是二十世纪的新生科学。至于分子遗传学到二十世纪下半叶才成为一门独立的学科。分子遗传学的发展，与化学、进化学说及遗传学所取得的成就是分不开的。本文试图主要从分子遗传学的角度，向读者介绍各门学科的某些手段和思维方法。

在遗传学中基因概念的形成，应直接追溯到孟德尔的工作。孟德尔的著作发表于 1865 年，但是直到 1900 年才得到科学界的公认。孟德尔阐明的中心思想是：动植物的各个特性由特定的基因所决定，而且基因在世代"相传"的过程中，可以观察到为一个个独立的实体。当然，孟德尔观察基因从上代到下代表露的唯一方法是在已知亲本特征的后代中，观察其外表特性。孟德尔通过豌豆的七对相对性状的遗传规律的研究，提出了遗传因子的概念，这七对相对性状是：高茎对矮茎，绿色种子对黄色种子，皱粒种子对饱满种子等(图 15)。

二十世纪上半叶，遗传方式的分析已经扩大到了动植物，包括人类在内的许多种属，并广泛地涉及到各种各样的特征特性。随着遗传分析向纵深发展，基因的概念也逐渐演变为以某种方式表达为个体的形态和生理特征的遗传信息。这一基本概念向二方面展开：(1)对基因本身最终进行化学分析，并揭示遗传信息如何编码，以及密码又如何在表型特征的发育过程中译读出来。(2)用历史进



表现型	遗传型	遗传型频率	表现型频率
高茎	*DD*	1	3
	Dd	2	
矮茎	*dd*	1	1

图 15　基因概念的实验基础

用符号表示如图的孟德尔的经典实验，在此实验的基础上产生了基因概念，在孟德尔以前，大多数生物学家都把遗传理解成某种融合现象，认为，后代表现为两个不同表型亲本的中间类型就是融合遗传的明证。然而，孟德尔通过一次研究一个特性的遗传模式，证明了经过许多世代后，表现型及决定表现型的因子(现称为基因)始终稳定不变。

化的观点来看基因概念，将之理解为在进化的历史长河中遗传信息的形成与传递，这就是时间矢量上的信息流问题。

现已知道，基因是由 DNA 组成的。编码在 DNA 分子上的信息被转录成 RNA 分子，并由这类 RNA 分子转译为蛋白质分子。

已知蛋白质分子在生物体和细胞中有着重要的作用，每一种蛋白质具有一种特定的功能。分子遗传学的内容就是研究每一种信息的大分子在行使特定功能方面的化学特性（图 8 和图 9）。在研究进化的过程中，人们颇感兴趣的是，这些信息大分子究竟如何演化成为它们象在今天这样的结构和功能。在现存的动植物和原核类生物物种中，信息大分子是通过其他的大分子得到信息，如果说现在的生物是从无生命的物质演化而来的话，那么，很显然信息的转移也并非从来就是如此。因为在有生命出现之前，一些分子通过前生命的物质进化过程慢慢地演化成多少有点复杂的物质，然后，当这些分子的数目和复杂性到达一定程度时，生物化学反应即便接踵而起，在这里有些反应能导致信息大分子的形成与发展，而信息大分子本身则又会反过来促进新的生化反应。

由大分子化学反应产生的一个饶有兴趣的问题是：信息如何贮存在基因(DNA)之中，以及化学的信息如何最终又表达为使之有别于别的有机体和细胞的表型。这一问题，如何从进化与化学两方面给予解释，必定会涉及到生物科学的许多方面——对此，决非一本书所讲得清楚，但是，我们可以用一个总的命题，把生物学系统内的信息流来作为这一研究的主题。因为从广义上来说，生物学信息是联结过去到现在和上代到下代的一条纽带，从狭义上来看，在细胞和有机体内的各种原件执行对生命所必须的许多职能是受生物学信息所指挥。总而言之，分子遗传学之所以引人入胜，只因为是科学正面临着这一挑战：如何把生物学语言翻译成普通的语言。

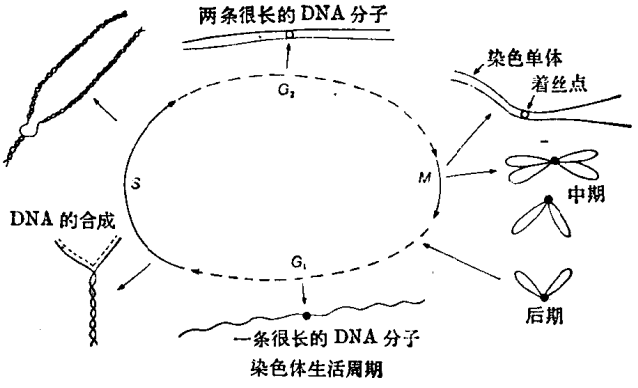
自从遗传学诞生的那天起，就把生物的信息流作为它的主要研究目标。现在不妨让我们按照发展顺序简要地回顾一下遗传学的历史。从孟德尔的首篇著作发表到人们认识和确认他的工作之间，植物学家和动物学家就相当详细地研究和描绘了被称为染色体的极小的细胞器。按照两种根本不同的细胞分裂类型，现已能确切地描述出染色体的形态和行为。第一种类型，称为有丝分裂，有时也称为细胞的均等分裂。由一个细胞产生出两个细胞，而且

每个子细胞和母细胞完全相同，也就是说，子细胞所含有的染色体数目和类型与母细胞完全一样，这是因为每一个染色体能正确地自我复制，并且复制后每对染色体中的一个进入到一个子细胞中(图 16a)。

由第二种类型的细胞分裂所产生的子细胞，它的形态和染色体数目不同于原初母细胞。这种分裂类型称为减数分裂，伴随着减数分裂形成了性细胞或配子，在高等动植物中，称作为精细胞和卵细胞的配子其染色体数目精确地是它们母细胞的一半。成熟分裂或经常被称为减数分裂的发现，导致了确认产生配子的母细胞具有两套染色体，并且由于减数分裂的结果使两套染色体分开，进入每一配子的只有一套染色体(图 16 b、c)。

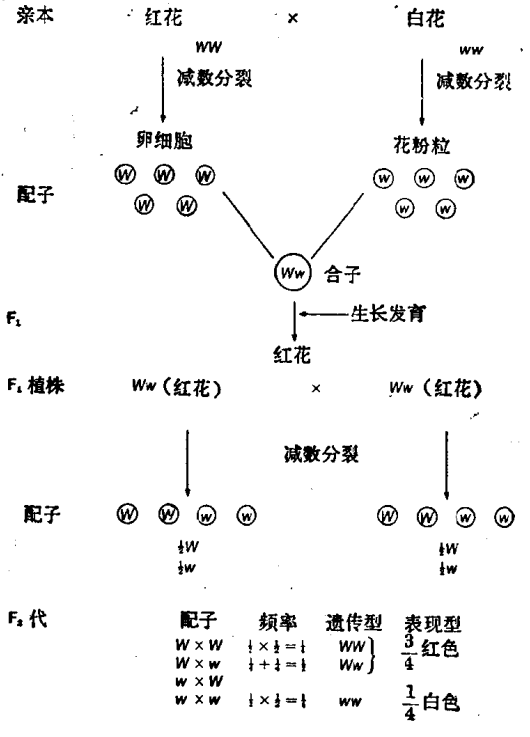
当孟德尔的工作发表两年后，有两个遗传学家分别发现孟德尔的遗传模式和染色体随机对分到配子中仅是同一现象的不同表现形式(图 17)。他们指出，孟德尔所说的因子，事实上存在于染色体上。当时对因子(基因)究竟是个什么东西并不知道，但是，想象中必定有某种因素会经由配子以某种方式从上代传递到下代，并且随之即伴同有相对性状非此即彼的产生，如植物的高和矮等，总而言之，当人们还不知道基因究竟是什么的时候，就提出基因定位在染色体上，这确实是一件了不起的事情。基因和染色体间这一奥妙的相互关系，就是通常所说的孟德尔遗传方式的实质。

在基因和染色体之间的对应关系被人们证实以前，很明显，当时所理解的孟德尔式的基因并不是真正的实体，而是抽象的概念或者是说明性的符号。这主要由于基因的存在，仅仅是在通过后代个体中表现出的相对性状才能证明。如图 16(b)所示：红花品种与白花品种杂交，所有后代植株均开红花，但是，当后代植株相互杂交，其结果一些后代开红花，另一些开白花。换言之，杂种第一代开红花的植株除了带有决定开红花的基因之外，还带有决定白花的基因。因此，某个 F_1 代的植株所携带的那些基因，通过该 F_1 植株的表型是不能确切识别出来的，只有通过对该植株后代表型的观察才能识别出该植株所携着的是何种基因。

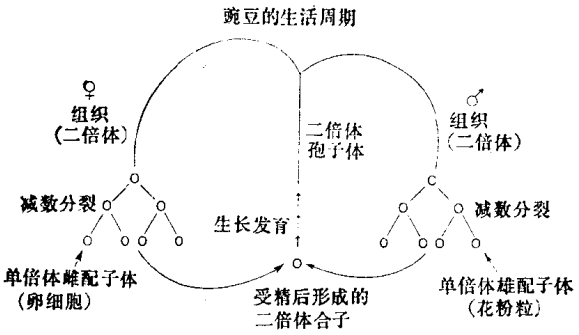

两条很长的DNA分子
染色单体
着丝点
G_2
S
M
中期
DNA 的合成
G_1
后期
一条很长的 DNA 分子

染色体生活周期

(a)


亲本
红花
×
白花
WW
ww
减数分裂
减数分裂
卵细胞
花粉粒
配子
合子
F_1
生长发育
红花
F_1 植株
Ww（红花）
×
Ww（红花）
减数分裂
配子
½W
½w
½W
½w
F_2 代
配子
频率
遗传型
表现型
W × W
W × w
w × W
w × w
½ × ½ = ¼
¼ + ¼ = ½
½ × ½ = ¼
WW
Ww
ww
¾ 红色
¼ 白色

(b)



(c)

图 16　染色体的生活周期和配子形成

(a) 此图表示染色体生活周期的大致情况。如果我们以M期(有丝分裂)开始,并按顺时针方向来观察有丝分裂的周期,那末我们将会发现所有染色体均是由两条细长的同源染色单体组成。在有丝分裂的始期(前期),染色单体缩短变粗,到分裂中期缩短至最大限度,到分裂后期(此时每条染色单体已变成为染色体)两条染色体分开,其中一条向细胞的一边移动,而另一条向相反的一边移动,以后就形成了二个新细胞。从其中的一个细胞来观察,可以看到往长向伸展的染色体,最终会变成一条很长的 DNA 分子链,这就是 G_1 期(生长期1)。到 S 期(DNA 合成时期),DNA 分子自我复制,到 S 末期复制成二条 DNA 分子长链,到 G_2 期经折褶和浓缩产生出两条染色单体,与M期开始时所见的相同。

(b) 与图 15 相似,用基因符号来表示染色体(W 为红色 w 为白色),二倍体细胞具有二套染色体,配子只具有一套染色体,当雌雄配子受精结合成合子,染色体数便恢复到二倍体,每个合子通过细胞的生长和发育产生出成熟植株。杂合体(F_1)植株产生出两种类型的配子,假定 F_1 的双亲产生出的两种配子的数目相等,同时受精作用的机会又均等,那么 F_2 代出现的遗传型和表现型的频率就象图(b)所表现的那样。此图大体上表明了所有高等动植物的生活周期。

(c) 豌豆的生活周期与另一些生物例如人类的生活周期的不同之处,主要表现在豌豆的雌雄性细胞产生于同一个个体。

上述推论是假定基因和染色体一样成对存在,并且每一对基因中的一个成员被定位在一对染色体的一条之上。因此,为研究遗传模式,不管什么种,要求亲本群体中的个体至少必须有一对不同的相对性状,也就是说,要证明基因从上代传递到下代至少必须要有一对基因作为鉴定标记。例如,在豌豆原始群体中,如果孟德尔没有发现高茎类型和矮茎类型的植株,那末他就不可能研究决定株高基因的遗传。我们假定高茎豌豆为"正常"类型,在正常类型的大群体中存在着少数"非正常"类型,即矮茎变异类型。用大

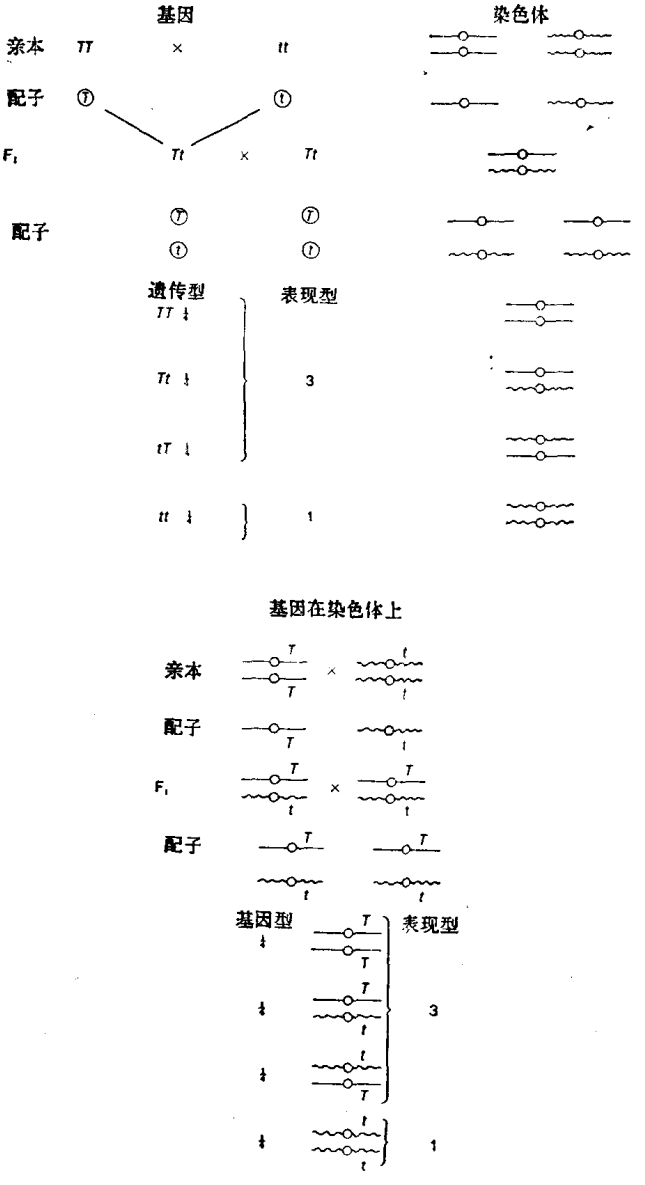


图 17　基因位于染色体上

通过基因分离模式的分析（通过表型观察）和在形成配子的过程中对染色体的分析（通过镜检），就得出了基因位于染色体上的结论（1902 年）。

写字母 T 表示决定高茎的基因，用小写字母 t 表示决定矮茎的突变基因，T 和 t 表示相对应的等位基因，或称一对等位基因。

在十九世纪下半叶，当植物学家和动物学家开始描绘染色体的时候，核酸(DNA 和 RNA)已被分离了出来，当时分离出的核酸是 DNA 和 RNA 的混合物，并未发现二者之间的差别。直至多年以后，这二类核酸分子的功能甚至还仍然只处于大体上可揣想的水平。到 1940 年，生物化学家和细胞生物学家发现大多数 DNA 存在于细胞核中，染色体的主要成分是 DNA，RNA 主要存在于细胞质中，而且旺盛的蛋白质合成(图 18)是细胞生长的显著特征。所以从 1940 年起，人们才开始认识到这类分子的重要性。

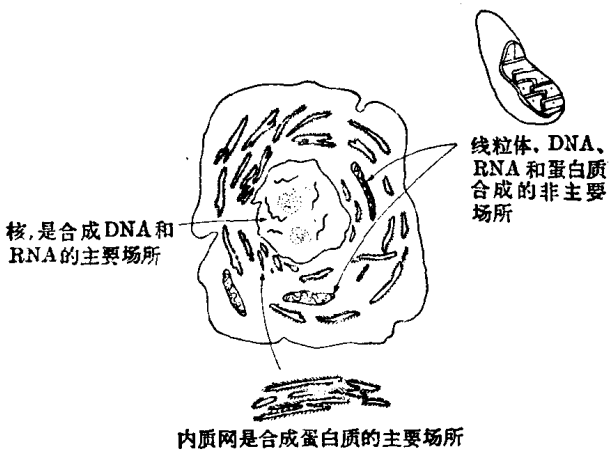


图 18　合成大分子的位置

DNA 的合成和累积主要是在细胞核中进行，RNA 也是细胞核中合成，但 RNA 主要存在于细胞质中。蛋白质在细胞中到处都有存在，但是它们是在细胞质中的核糖体上合成的，核糖体绝大部分是存在于内质网(细胞质膜)上。

生物学的一个分支即生化遗传学对遗传学的发展直接起过转折性的作用。为了便于人们了解生化遗传学的基本含义，读者不妨再回忆一下前面所举的那张人体生活周期图（图 1)。每个人无论是男性还是女性，都是由受精卵或合子发展而来的，在合子中包含有他或她一生赋有的全套遗传信息，通过细胞分裂和分化，

合子才得以生长，随着这一过程的持续进行个体的表型特征才会出现和发育。发育受遗传信息所支配，但是环境也给予一定的影响，决定眼睛颜色、毛发颜色、个体高度等等基因，到表达出这些基因的功能要很长时间。也就是说，在决定颜色、高度等基因"作用"下，从这些性状开始发育时起，到表型特征的出现要相隔很长一段时间。同时，千百万个其他基因在个体发育中，必定会行使它们各自独特的职能。由于这一缘故，就迫使人们去寻找可以直接观察基因"作用"和表型发育之间的关系，生化遗传学就是一门特定的对遗传背景直接进行分析的学科。

在某种程度上，生化遗传的历史与遗传学的历史颇相类似，第一篇关于生化遗传工作的文章与遗传学的首篇著作发表以后的遭遇一样，在相当一阶段时间内，也没有得到科学界的普遍尝识。在1909年，英国的一个医生 Archibald Garrod 出版了一本书，名叫《先天性代谢病》(Inborn Errors of Metabolism)。在这本书中，Garrod 尽可能通过相当数量的世代研究了某些罕见的代谢病的遗传模式，同时，他也注意到"病员"的全部表型与特殊的代谢病之间的联系（图 19)。必须指出，当时人们对于新陈代谢的知识很浅，关于酶实质上还不知道，至于酶和基因的关系同样一无所知。即使如此，Garrod 却能正确地确定特殊的代谢缺陷与突变基因之间是一对一的关系，也就是说，他第一个指出了代谢中出现的错误，为存在有突变了的基因提供了表型上的肉眼可见的证据，但是至于这同属一个基因的两种形式(等位基因)是如何会把性状发育引向两种绝然不同的结果，究竟原因何在？作者本人没有回答，这个问题，经历了三十多年以后，才由其他科学家最终加以阐明(图 20)。

在 1909～1940 年间，几个遗传学家跟踪追击基因和蛋白质之间的相互关系。例如，后来成为遗传学的一个分支即群体遗传学的创始人之一 Sewall Wright，在 1917 年提出决定豚鼠毛皮颜色的基因启动一系列特殊的化学反应，从而使豚鼠具有一定的毛色。但是直到三十年代中期，对于基因和表型之间的对应关系几乎没有展开工作。1938 年，另一个群体遗传学的创始人 J. B. S.

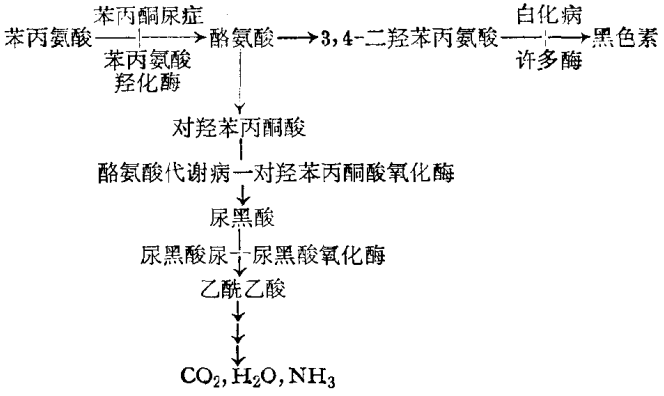


图 19　人的先天性代谢病

在本世纪的头十年中，Archibald Garrod 首先描述了先天性代谢病。当时Garrod 还不知道图中所示的有关酶的情况，但是他已把下列代谢病与孟德尔的遗传原理联系起来研究。

苯丙酮尿症：在血液和尿中，苯丙氨酸及其有关的化合物积累起来，使中枢神经系统受损，因此，引起智力发育受阻成为白痴。通过验尿，发现有苯丙酮酸的存在，便可确认为是苯丙酮尿症患者。

白化病：白化病患者缺乏黑色素，因此病员头发呈白色。小时候的眼睛表现出红色和彩桃红色，其皮肤呈粉红色。因为黑色素是由其前体酪氨酸通过某些酶的催化反应才合成的，所以如果在上述反应中某些基因产生突变，就会导致产生白化病。

酪氨酸代谢病：关于这一综合症，人们知道很少，部分原因是患此病的人十分罕见，但更主要的是这种病人的特征类似于苯丙酮尿症患者。

黑尿病：当缺乏尿黑酸氧化酶时，在体内就会积累尿黑酸，并且在尿中可以测定出来。当含有尿黑酸的尿被暴露在空气中，也就会转变成黑色。某些老年患者最终将发展为软骨色素沉淀和关节炎。

Haldane 阐明了称为抗原物质与基因之间表现为精确的一对一关系的这一事实(图 21)。例如，关于 AOB 血型，A 型血的人，仅具有A的抗原，B型血的人，仅具有B的抗原，AB型血的人，具有A和B两种抗原，O型血的人，既没有A抗原，也没有B抗原。血型作为表现型，因此通过它的不同特征把人群在血型上彼此区别开来。若用一种基因符号来表示与之相对应的一种表型，那末人的 AOB 血型的等位基因分别是：I^A, I^B 和 I°，A血型的基因型为 I^AI^A 或 I^AI°，AB 血型的基因型为 I^AI^B 等等。

Haldane 着重指出了在高等动植物的细胞和器官中发现的各种各样的蛋白质，在功能上有着高度的专一性。从细胞和器官中

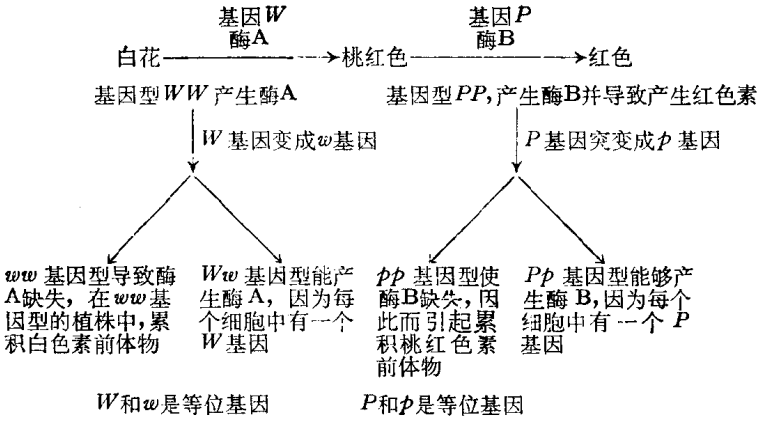
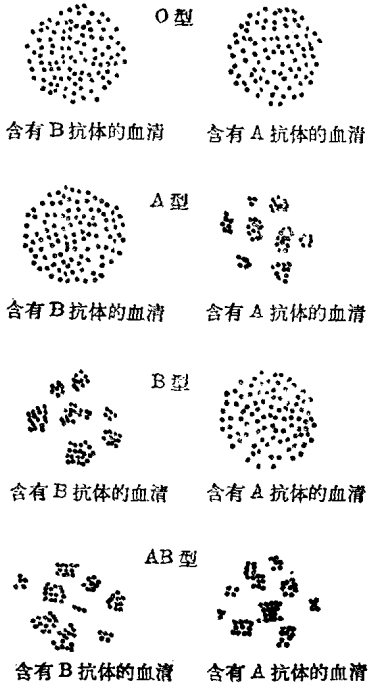


图 20　突变产生不同表型性状的机制

此图说明花色变异的突变机制。形成红色素致少有两个前体物，即白色素和桃红色素。白色素通过A酶的酶促反应转化成桃红色素，并由桃红色素，通过B酶的酶促反应转化成红色素。因此如果A酶失活，那么产生白花的白色素前体就累积起来，将来便开白花，如果A酶能起作用，而B酶失活，那么就积累桃红色素前体。杂合的植株表型之所以正常，只因为是每个细胞中各有一个 A、B 显性基因便足够起作用的缘故。

所发生的千百万个代谢反应来说，这种专一性尤为明显，每一种代谢反应均要有特定的酶来催化。Haldane 又指出，蛋白质功能的专一性是由它的结构的特异性所决定，而且在某些方面，每一种蛋白质的结构是由特定的基因所决定。

当 Haldane 推测关于酶、抗原和基因之间相互关系的时候，在巴黎的 Boris Ephrussi 实验室中，正在进行研究有关这方面的颇有价值的实验工作。同 Ephrussi 一道工作的有美国遗传学家 George Beadle。Ephrussi 和 Beadle 通过试验试图弄清楚果蝇 (*Drosophila melanogaster*)的眼睛色素的遗传控制(图 22)。他们的实验依据是基于果蝇眼睛颜色的每一种色素，按照一定的反应顺序合成，而且每一个合成步骤均需通过特定的酶促作用才得以实现。认为基因突变会导致酶的缺失，即可据以来解释果蝇眼睛颜色不同的现象。由于某一酶的缺失就会使某个代谢途径阻断，从而代谢途径阻断之前的色素前体物被累积，使最终的色素便不能形成。Ephrussi 和 Beadle 通过观察大量的不同眼睛颜色的突



凝集反应

血清中的抗体类型	血型			
	O	A	B	AB
A抗体	No	Yes	No	Yes
B抗体	No	No	Yes	Yes

遗传型	表现型
$I^{A_1}I^{A_1}$	A_1
$I^{A_1}I^{A_2}$	
$I^{A_1}I^{O}$	
$I^{A_2}I^{A_2}$	A_2
$I^{A_2}I^{O}$	
$I^{B}I^{B}$	B
$I^{B}I^{O}$	
$I^{A_1}I^{B}$	A_1B
$I^{A_2}I^{B}$	A_2B
$I^{O}I^{O}$	O

图 21　基因和血红蛋白

人群可按血红蛋白的不同类型如A、O、B加以区分，到目前为止已知道有六种表现型：A_1、A_2、B、A_1B、A_2B和O。产生上述表现型的基因型如上图所示，在图中表示出每对等位基因和特定的蛋白质之间有着1:1的相互关系。在图例中，A抗体和B抗体的血清同四种血型A、B、AB和O的血分别混合，如果出现凝集现象，那么就表示出与细胞红血蛋白起凝血反应。

变果蝇后，描绘出果蝇眼睛色素代谢的一般概况。但是，(对于原因的解释) 他们没有能够证明突变基因和突变酶之间一对一的相对应的关系。

Beadle 返回美国以后，对果蝇眼睛色素的生化研究又继续工作了四年，方使他本人完全相信，为了证明一个基因决定一个酶的假设，必须利用比较简单的生物作为研究对象。果蝇虽然比大象简单，但是从遗传的角度来看，果蝇仍然是一种高度复杂的生物。

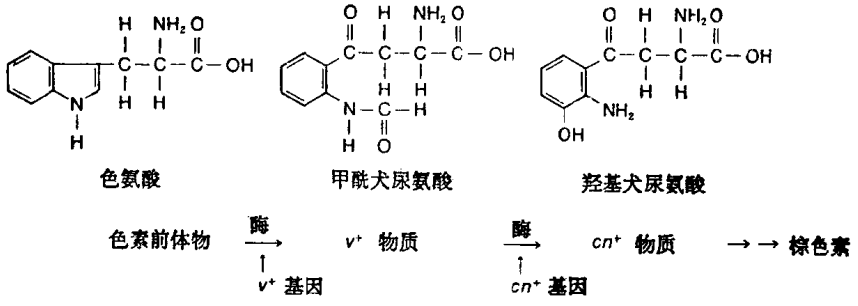


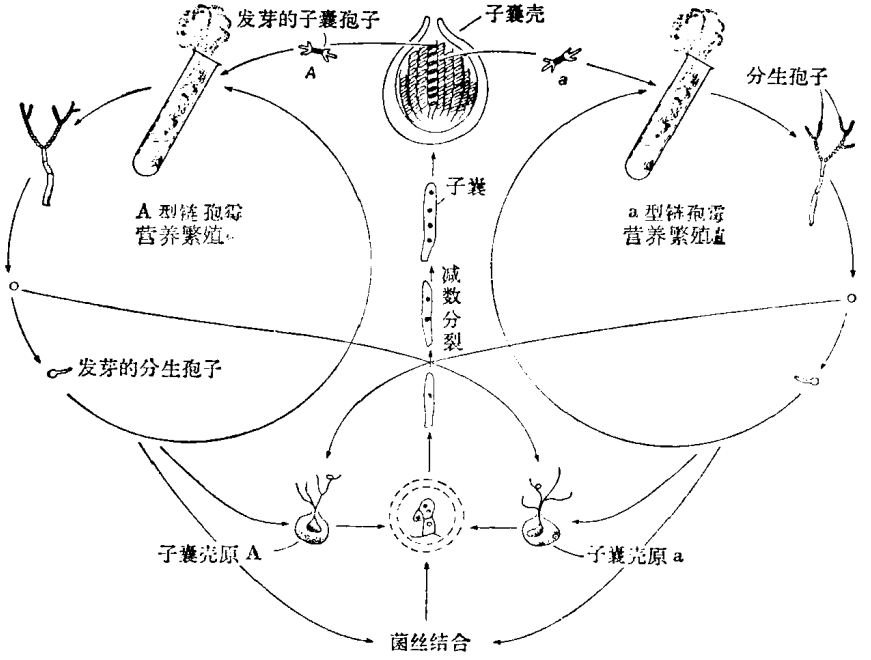
图 22　果蝇的眼睛颜色突变

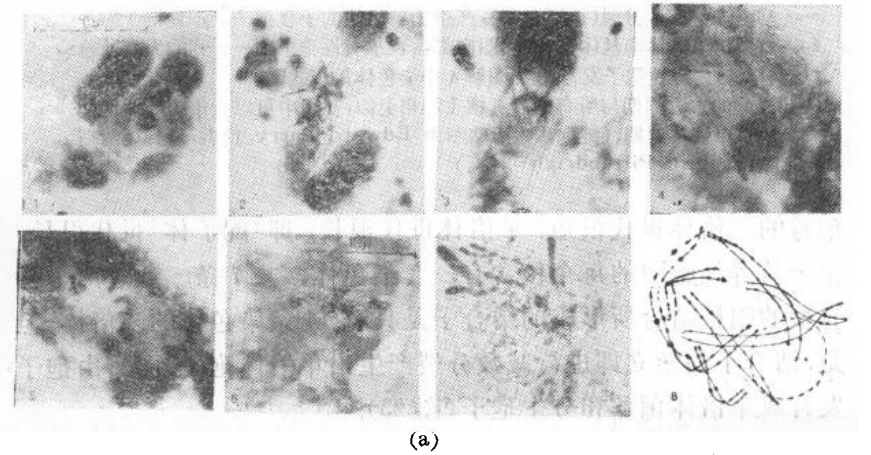
色氨酸是果蝇眼睛颜色色素的早期前体物。这一前体物通过 v^+基因控制酶促反应产生 v^+物质。v^+物质通过 cn^+基因控制酶促反应，使 v^+ 物质转化成 cn^+ 物质。v^+物质是甲酰犬尿氨酸，cn^+物质是羟基犬尿氨酸。

因为果蝇和大象以及人体一样，在每一个细胞中具有两套染色体(二倍体染色体数)。在任何一个二倍体生物中，每对基因有二个等位基因，由此必定会使遗传分析带来复杂化，例如，假使一个等位基因是显性，则另一个是隐性，由于显性掩盖隐性的缘故，隐性基因的作用就得不到表现；又如果每个等位基因都发挥作用，那么很难把等位基因的两个成员的作用区分开来。因此 Beadle 选定只具有一套染色体(即单倍体)的生物作为遗传研究的材料，就避免了显性掩盖隐性的现象以及避免了等位基因之间的相互作用等等的问题。

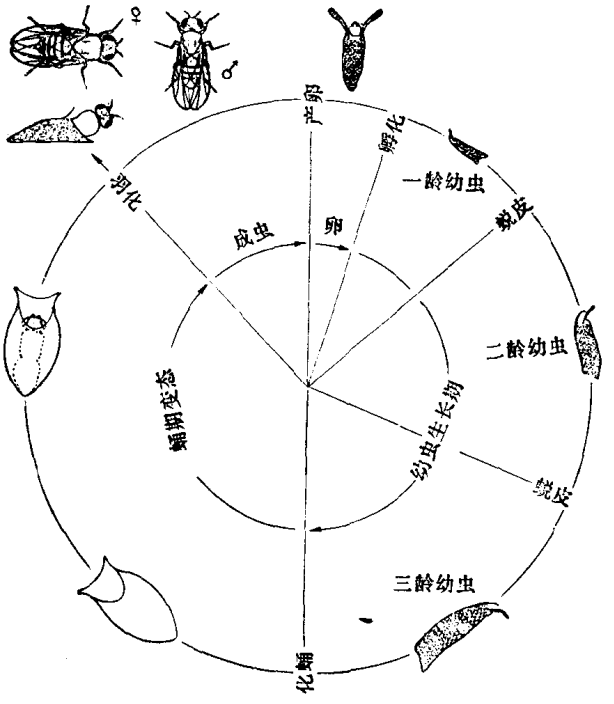
Beadle 在研究色素形成的生化工作中，提出了利用单倍体的建设性意见，这对于生化学家的实验构思、实施以及对实验结果的分析解释极为重要。遗传学家 Beadle 的工作，生物化学家 Edward Tatum 的工作，及单倍体粗糙链孢霉 (*Neurospora crassa*) 的工作，到 1940 年，这三个方面成为有机的统一整体。Beadle 和 Tatum 通过实验证明了某些代谢途径的阻断与某些突变基因之间存在着一对一的对应关系；此外，他们第一个架起连接遗传学和生物化学的桥梁，从而揭开了通向研究基因功能的新的实验途径，因此，当人们翻开生化遗传学的历史，参照他们的工作和设想，很自然地把他们称之为分子遗传学的奠基人，那是理所当然的。

链孢霉比病毒复杂，但比果蝇简单(图 23)。对此，只要通过图 23 比较它们的生活周期就一目了然。果蝇从受精卵开始直至死亡是一个二倍体有机体，它们产生的精卵细胞才是单倍体。链





(a)



(b)

图 23 链孢霉和果蝇的生活周期

(a) 在链孢霉的生活周期中，通过断裂或分生作用的无性繁殖方式产生出单倍体菌丝。有性繁殖需要有二个接合型（*A* 和 *a*）。二个接合型的核结合后，马上进行减数分裂（还原分裂），产生出单倍体的子囊孢子。子囊孢子发芽产生出单倍体菌丝体。在(*a*)图底部，表示有性孢子的一系列的发育阶段(1～7)为有性孢子发育，8 为图解 4 中染色体的模式图。

(b) 果蝇的生活周期与所有高等真核生物相类似，在图中可以看出二倍体时期超过单倍体时期。(链孢霉生活期图由 Edward Barry 提供，果蝇生活期图由 Colin Pittendrigh 提供。)

孢霉的二倍体世代很短，单倍体世代很长，即"配子体"世代很长。位于营养细胞中的每个核含有一套染色体，共 7 条。通过两个接合型的胞核结合所形成的结合子是具有 14 条染色体的二倍体，但是，结合子的核立即进行减数分裂产生出单倍体孢子，而后由孢子发育成单倍体菌丝和分生孢子(图 23)。

通过上述生活周期的比较，说明链孢霉比果蝇简单。另外，按

照自养程度和异养程度的高低同样说明链孢霉比果蝇简单，一个自养有机体能够从简单的无机分子(水、氨和一定的矿物元素)和二氧化碳合成对于维持生命所必需的非常复杂的有机分子。大多数植物和许多原核类生物能够从这些简单的前体物合成氨基酸、维生素、脂类以及其他一些不可缺少的物质。这就是说，大多数植物和许多原核类生物有着很多种酶，有了这些酶才能把无机前体分子合成为维持生命所必需的有机分子。

异养型生物，如果蝇，缺少这类酶，因此，它必须从提供的食物中获得一些所需的有机分子。又如人体，组成蛋白质的 20 种氨基酸中有 8 到 9 种不能自身合成。然而，链孢霉能够合成所有的 20 种氨基酸(图 24)。所以人的食物中，除了有多种维生素及其他有机分子之外，还需要含有维持生命所必须的 8 到 9 种氨基酸。

一般来说，愈是复杂的种类，异养程度就愈高(复杂的是指与人关系较近的生物，简单的是指与细菌和病毒关系较近的生物)。

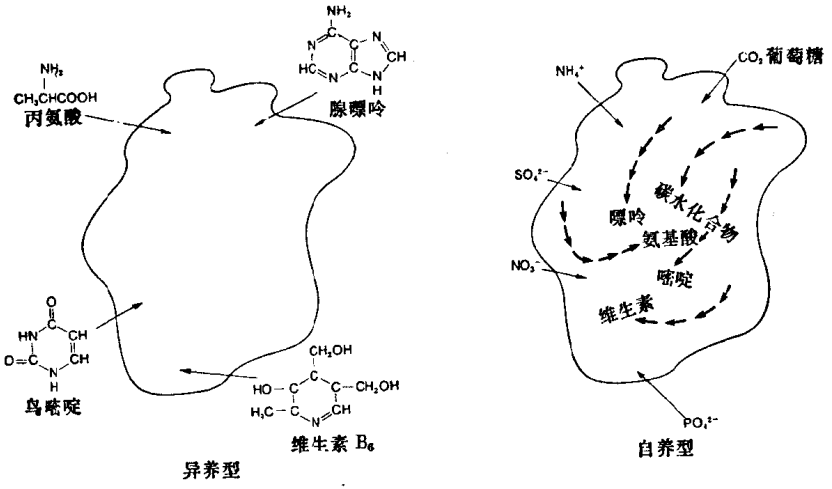


图 24　异养型和自养型

异养型是指依靠外界供给一种或多种必需的养料方能生存；自养型是指不需要依靠外界供给养料也能生活。自养型能够从无机离子(包括H.C.O.N.P 离子) 和微量元素合成必需的养料。因而能够独立生存。此外自养型生物还需要提供碳源如葡萄糖。人和果蝇是属于异养型，所以需要供给他们多种氨基酸、维生素和其他必需的养料。

然而，在整个生物世界中，完全自养和高度异养之间有一个广泛的变异幅度。因此在实际使用这两个术语时，并不严格体现字面本身原先的确切含义而只是表示依靠合成前的有机分子的程度。

根据上述定义，可知链孢霉并不是完全自养的，因为需要供给它一种维生素——生物素方能生长。培养在含有一定的矿物质、无机盐、生物素和糖的培养基上的链孢霉细胞，能够合成所有其余的维生素，全部 20 种氨基酸和作为构成细胞脂类分子所需的脂肪酸，以及能够合成大分子 DNA 和 RNA 的建筑元件嘌呤和嘧啶。这意味着链孢霉之所以具有合成上述物质的能力，维持其生长，至关重要的是因为链孢霉细胞有着许多种酶。这一认识正是 Beadle 和 Tatum 赖以进行实验设计的依据。

Beadle 和 Tatum 利用异养性较高的链孢霉为材料，通过理

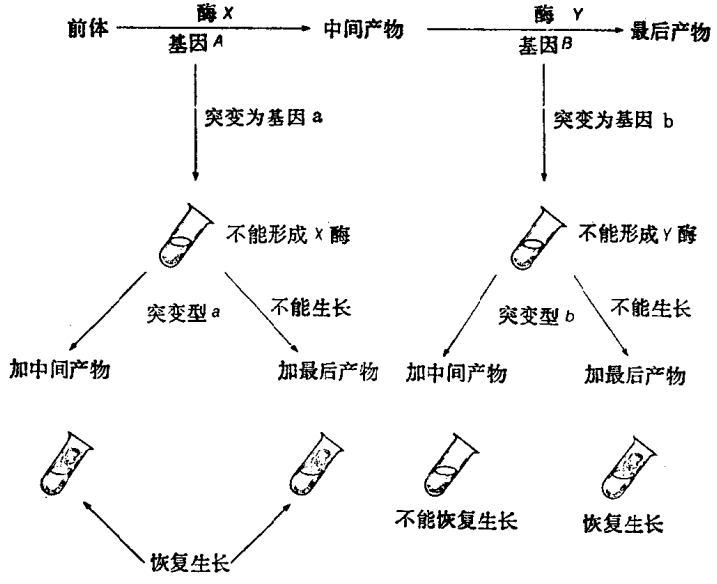


图 25　链孢霉的生化突变

通过X射线等处理，诱发基因突变，致使催化至关重要的代谢反应的酶失活，在这种情况下，如果不及时给有机体提供所缺少的中间产物，它就不能生长，成为致死突变。某一基因的突变导致某一酶的缺失，这种一对一的相互关系，在 Beadle 和 Tatum 的实验中，得到了科学的证明，后来诱发基因突变的方法也被人们用来研究中间产物的代谢途径。

化因素诱发基因突变的方法，研究基因和酶的关系。他们的实验是这样的：先用 X 射线、紫外光和一定的化学物质诱发突变，然后在不加入外源物质的情况下进行培养，从中筛选出许多除非提供特定的代谢产物，否则便不能生长的突变品系(图 25)。这些突变品系的表型特征是当缺少特殊的氨基酸、维生素、嘌呤和嘧啶时，就不能生长。一般来说，这些突变品系遗传方式合乎孟德尔规律，多数突变品系总是缺少一个酶。这样研究一个基因一个酶的著名假说，才终于由 Beadle 和 Tatum 借助于诱发基因突变的研究，取得了科学界的公认，三十多年来，这一理论除了在某些细微末节上稍有修正外，直到今天仍然巍然不动。

基因和基因作用

在详细论述研究基因作用的生化遗传以前，为了说明基因概念的产生和发展，让我们简要地回顾一下它们所走的里程。如前所述，基因这一概念起始于孟德尔的工作，但是正式建立基因学说，并得到迅速发展乃是本世纪前半叶的事。

在遗传研究上，孟德尔和他的前辈的不同之处在于他一次研究一个或几个相对性状的遗传情况。在孟德尔之前，遗传性状从亲代传递到子代被看成是一种融合过程，即子代的性状呈现双亲的中间表型，甚至达尔文也以这种混合格式对亲代和子代加以比较，在达尔文看来，性状在世代之间的传递仿佛是一种掺和效应。但是，孟德尔连续多代观察了表型的单个性状，从未发现有任何性状的掺和或更改。

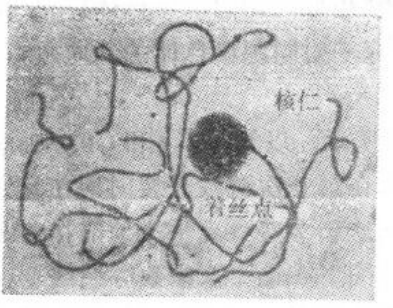
孟德尔所研究的由单因子决定的相对性状如豌豆植株的高与矮是由一个基因的相对形式，即所谓的等位基因所决定(豌豆高度与人体高度在遗传上不能混为一谈，因为人体高度是由多基因决定的)。一般把高茎豌豆的基因用大写字母 T 表示，矮茎豌豆的基因用小写字母 t 表示。在二倍体豌豆植株中必定会携带有二个决定高度的基因，高茎豌豆中的基因为 TT 或 Tt，因为 T 对 t 是显性，所以携带 Tt 基因的豌豆植株仍然表现高茎。矮茎植株中的基因则

为 tt。将基因为 TT 的高茎植株相互杂交，只出现 TT 基因的高茎后代；将植株基因为 tt 的矮茎植株相互杂交，仅仅出现 tt 基因的矮茎后代。然而颇有意思的是将基因为 Tt 的杂合体植株相互杂交，其子代既有高茎类型，也有矮茎类型。并且高茎与矮茎之比是 3:1(图 15)。孟德尔虽未明确提出信息流的概念，但是他用符号 T 和 t 来暗示颗粒遗传的含义，T 表示高茎信息，t 表示矮茎信息。

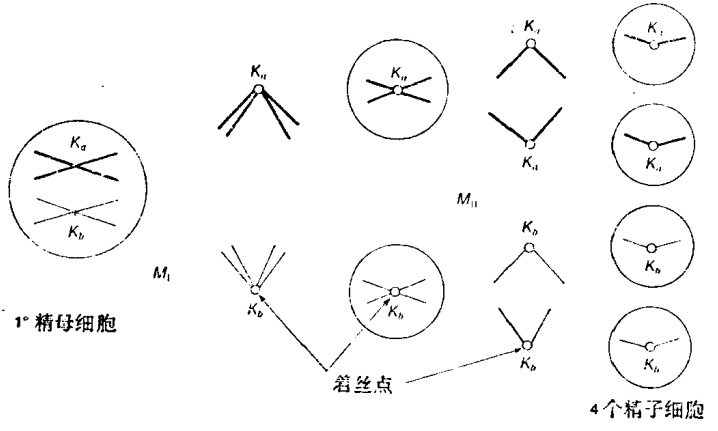
孟德尔对遗传理论的主要贡献，在于他通过他的实验资料的分析综合，从而归纳出一个简单的遗传模式，也就是后来人们称之为孟德尔的遗传定律。我们将在下面对分离和重组这两个定律作进一步的讨论（孟德尔的独立分配定律后来被人们称为重组定律)。现在我们知道，可以把这两个定律应用到来自于单倍体配子体相结合而产生的所有二倍体生物。(在单倍体生物中，同样发现有分离和重组现象，但尚未发现如同孟德尔在二倍体生物中所描述的那种遗传模型。)

分离可以看作在配子形成的过程中等位基因的相互拆离，重组可以看作是在配子形成的过程中非等位基因的自由搭配。因为二倍体有机体具有两套完整的染色体，所以一对同源染色体上有一对等位基因的二个成员，不是同质基因 (TT)，就是异质基因 (Tt)。每个配子只具有一套完整的染色体，所以每个配子只有一对等位基因中的一个成员，T 或 t，决不会两个成员同时存在。事实上，配子形成的研究使得孟德尔的分离定律有了理论根据，也就是可以把同源染色体分别分配到配子中的事实与等位基因分离的现象视为同一过程的不同形式(图 17)。

一定种类的植物或动物都有着恒定的染色体数目，如在配子的核中(单倍体)发现豌豆的染色体数是 7 条，果蝇是 4 条，人类 23条。而在产生配子的细胞的细胞核中（产生卵细胞的初级卵母细胞，产生精子的初级精母细胞)具有两套染色体，人类总共有 46 条，果蝇 8 条等等。在每一套染色体之内的各个成员其形状、大小和着丝点部位通常是各不相同的（图 26)。根据染色体的不同外



(a)



(b)

图 26　孟德尔的分离定律

(a) 图中是玉米的染色体，可以看到其中有一条染色体与一个核仁相连，以及每条染色体各自的着丝点部位。(Corn chromosome from R.L. Philips in D. B. Walden(ed.), Genetics and Breeding of Maize wiley, New York, 1976.)

(b) 仅以一对染色体为例，说明一个小孢子母细胞产生出4个精子细胞的减数分裂模式图；K_a和K_b表示在第一次减数分裂时分离的同源着丝点。

形特征（形状大小和着丝点部位）给予编号，例如人体23条染色体，从大到小进行编号，规定最大的染色体其编号数最小，最小的染色体编号数最大。比较同一种的两套染色体，同一编号数的染色体形状相同，如一号染色体的形状相同，二号染色体的形状相同等等，如此类推。也就是说，同一个种的染色体，就每对而言，其形

状、大小等等极为"相似"(即同源染色体)。从减数分裂起到形成配子为止,通过研究一对同源染色体的动向,使得分离定律基本上成为肉眼可见。

如果象图 26b所示,以符号 Ka 和 Kb 分别表示二条同源染色体的着丝点,那末在第一次减数分裂时期在它们进入两个新细胞之前,我们会看到这二个着丝点平排地靠在一起。通常把着丝点彼此分开称作为分离。在第二次减数分裂时期,每个着丝点沿染色体长轴平行分裂,并分别进入到二个新细胞中。第二次减数分裂之后,产生出四个新细胞,每一个新细胞将发育成一个配子。(卵细胞的形成机制与此稍有不同,但在分离的概念上没有区别。)其中二个新细胞带有 Ka 着丝点,另外二个新细胞带有 Kb 着丝点。由同一种类的孢原母细胞所产生的配子的大群体中,有 50% 的配子带有 Ka 着丝点,另有 50% 的配子带有 Kb 着丝点。

假定把一对同源染色体的二个着丝点部位,看作是一对等位基因的位置,那末着丝点 Ka 就可视为一对等位基因中的一个成员,Kb 可视为另一个成员,显然可以用基因 T 代替着丝点 Ka,基因 t 代替着丝点 Kb,同样通过减数分裂的过程,最终产生出 50% 的配子含有 T 基因,另一个 50% 的配子含有 t 基因。由此可见,同源染色体的分开类同于等位基因的分离,正是染色体与基因这一平衡关系,才使得人们正确无误地得出基因在染色体上的这一推论。

当孟德尔同时研究二对或三对基因的遗传规律时,他发现非等位基因之间的分离彼此独立,互不干扰,由此建立了他的独立分配定律。今天我们把它称之为自由组合定律。现以二对同源染色体为例,从其在形成配子时,减数分裂中的动向来说明这个定律(图 27)。假定一次观察一对等位基因,那末在减数分裂时便可观察到每对同源染色体的"等位"着丝点的分开,是发生在第一次减数分裂时期。并且发现 1/2 的配子带有一种着丝点,另有 1/2 的配子带有另一种着丝点。但是,如果同时察看位于二对同源染色体上的二对等位基因的分离情况,那么就会观察到 1/4 的配子带

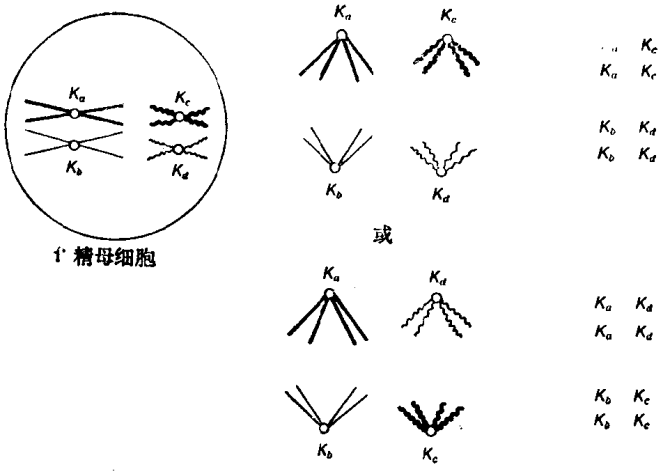


图 27　孟德尔的自由组合定律

位于非同源染色体上的基因，在配子发生时彼此独立分离。

有 *Ka* 和 *Kc* 的着丝点，1/4 的配子带有 *Ka* 和 *Kd* 着丝点。1/4 的配子带有 *Kb* 和 *Kc* 的着丝点，1/4 的配子带有 *Kb* 和 *Kd* 的着丝点，这就告诉我们，*Ka* 和 *Kc* 一起进入同一配子和 *Ka* 同 *Kd* 一起进入另一个配子，事实上这是一个随机的问题，即机会均等，同样，对于 *Kb* 和 *Kd*、*Kb* 和 *Kc* 来说也一样(图 27)。

虽然后人证明了孟德尔所研究的七对相对基因分别定位在七对染色体上，但是孟德尔在研究豌豆性状遗传规律时，并不知道这一事实。因此，孟德尔研究出豌豆的七对相对性状的遗传模式，正好碰巧彼此是独立分离的非等位基因。假如当时他知道豌豆有七对染色体，那末或许他会推测染色体就是基因。另外，一些遗传学家曾指出，如果孟德尔观察到的是非独立分离现象，其结果必定会使他束手无策，从而也无法归纳出遗传学上著名的两个基本定律。但是，不管怎么样，孟德尔毕竟还是给人们留下了一幅十分清晰而有重要价值的分离和自由组合的遗传模式图，在此基础上，当孟德尔的著作发表四十多年后，才开始形成和发展了基因的概念。

自由组合适用于非同源染色体和位于不同对染色体上的非等位基因。但是，当人们研究了动植物的所有种属时，会发现其染色

体数目大大少于基因的数目。因此，不难设想，在每对染色体上必定载有许多基因对，甚至有成百上千个基因对。可见，并非是所有的基因彼此都能够独立分离，事实也果真如此，在图 28 中，位于同一条染色体区段上的许多连锁基因当然不会独立分离。但是，上述连锁区段上的基因并非绝对不可分离，因为在减数分裂前期，同源染色体紧密配对的同时，会产生部分交换，也就是说，同染染色体间会发生某些基因的交换。

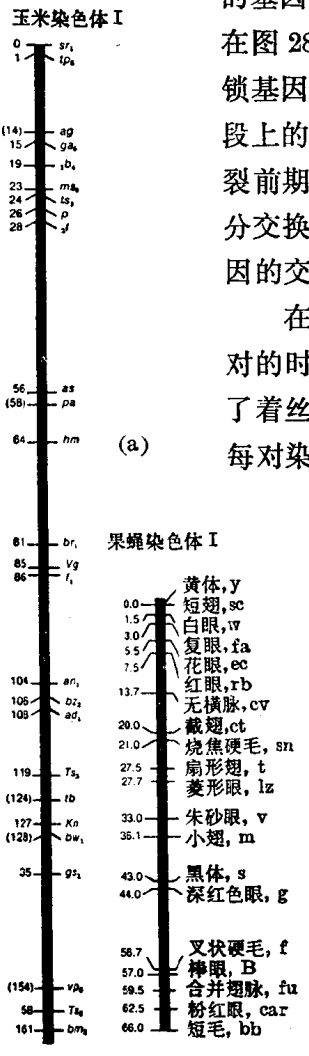


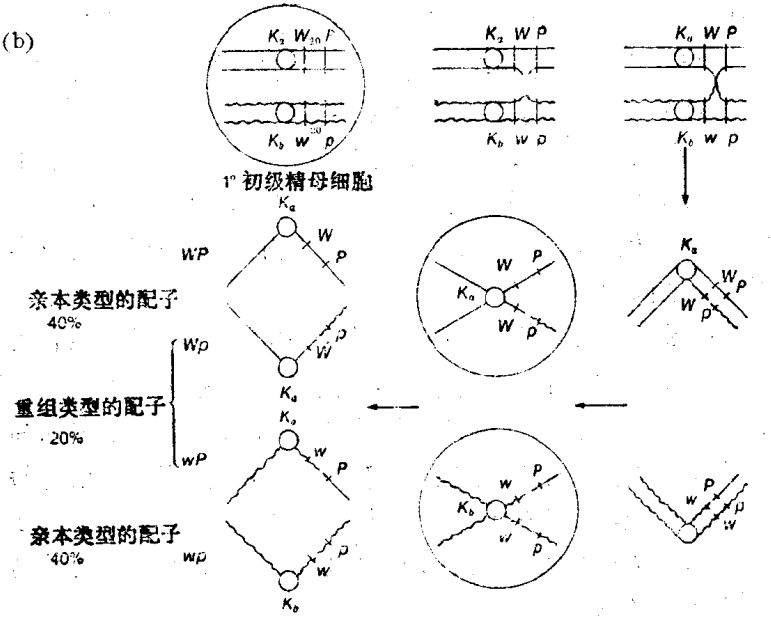
图 28　非独立分离和染色体图

(a) 染色体图反映染色体上成直线排列的基因位置，基因在染色体上的位置是通过重组频率来确定的。

在减数分裂的早期，同源染色体紧密地配对的时候，可以看到一对染色体的每个成员除了着丝点以外，已经完成加倍，也就是说，此时每对染色体是由 4 股或 4 条染色单体组成（图 28），亦即一条染色体，由 2 条染色单体组成。同源染色体之间的交换，发生交换的并非整个染色体，而只是染色单体之间发生交换亦即为半染色体交换。交换的部位是随机的，只是在靠近一个交换处要发生第二次交换，往往会受到阻抑。在一条很长的染色体上发生 4 到 6 次的交换是司空见惯的，染色体愈短，发生交换的次数亦随之减少。同源染色体之间的互换（交换）结果，便“形成”了父方和母方相对应的染色体片断交换了的新染色体（图 28）。

在整个染色体上的任何部位如果发生互换机率是随机的话，那末染色体上任何两点之间的距离可用此两点间的互换频率来表示，人们也正是根

(b)



W 与 w 分离，1:1
P 与 p 分离，1:1
} W 与 P 连锁，因此除非它们之间发生一次交换，总是一起分离。

Wp 与 wP 在整个群体中的频率等于 W、P 之间的图距。

图 28

(b) 如果基因彼此独立分离，那末后代出现重组类型的频率与出现亲本类型的频率相等(50%)。如果后代的重组频率低于 50%，则表明基因连锁在同一条染色体上。连锁基因之间的重组频率是随着基因的距离而变化的，而距离本身则又视基因间将来所发生的交换机率而定。例如：

独立分离 (W和P在不同的染色体上)	连锁
亲本 $WwPp \times wwpp$	$WwPp \times wwpp$
子代 $WwPp$, $wwpp$ } 亲本类型50%	$WwPp$, $wwpp$ } 80%
$Wwpp$, $wwPp$ } 重组类型50%	$Wwpp$, $wwPp$ } 20%

据这一标准来绘制染色体图的(图28)，从图28(b)上可以看出亲本类型WP/wp 和 wp/wp，这里的W和P是连锁在一条同源染色体上，所以后代决不会出现Wp/wp和wP/wp的新类型，除非在形成配子时，在W/w和P/p之间发生交换才会出现重组类型。如果W和P在一条染色体上紧密连锁，那末杂交后代的绝大部分个体是WP/wp 和 wp/wp，象亲本类型，小部分后代将是重组类型即Wp/wp和wP/wp，以及在后代总群体之内，重组后代的百分率表示W与P之间的图距。

这种图距单位与其他的测量单位如 μm 和 Å 不同，但它仍然是个度量单位，就是说，遗传学图是表示基因在染色体上的一维性座位量度，相当于几何学上的一把尺子，两个连锁较远的基因间的重组频率比两个紧密连锁的基因间的重组频率要高，其值(图距)就是沿着染色体长度方向的累加(假定A与B之间的距离为5个单位；B和C之间的距离也是5个单位，那么C到A或接近A的距离就等于10个图距单位)。

因此，重组有两种基本形式，即独立分配和非独立分配。如果任何两个非等位基因间的重组频率表示出非独立分离，那末可认为这两个基因必定连锁在同一条染色体上；然而，如果表现出独立分离，那末也未必一定能说明这两个基因位于不同的染色体上。因为重组百分率的幅度在0～50之间，位于同一条染色体上的基因其重组百分率可高达50%。因此，要确定两个相距较远的基因之间的确切关系唯一的办法只有设法去寻找位于这两个基因中间的其他一些基因，绘定它们的位置，以便最终来判定这两个基因是否位于同一条染色体上(图28)。

所有这些可归结为遗传图是一种说明概率的图表，是把假设的位点(基因通过表型才能观察到)以直线排列的方式按放到真实的线状染色体上的一种统计学方法。基因(在染色体图上的一个位点)是观察到的表型的符号。因而，不可能直接看到基因之间的重组，仅仅是通过特定的表型特征在子代中出现重组的频率来推知基因间重组的存在。可见由这类研究引伸出来的基因概念是抽象

的。因为，至少可以这样说，从来没有直接看到确确实实的有功能的基因。《基因论》(The Theory of the Gene)一书由 T.H. 摩尔根著于 1926 年，他扼要地总结了在本世纪头二十五年中所研究的基因学说的全部内容，同时，这本书主要论述关于基因的分离和重组。摩尔根雄辩地证明了关于基因和染色体之间的相互关系，这一相互关系的揭示，使得人们"相信"基因是真正客观存在的物质实体。

摩尔根例举了基因的许多与重组和分离无关的"属性"，其中之一就是基因决定着特定的表型特性，换言之，基因本身具有特定的重要职能。虽然重组和分离现象能告诉我们有关染色体的某些特性，但是它长期不能回答人们提出的如下问题：基因是什么？基因的功能又是什么？基因是如何行使它们的功能的？正因为这些问题推动了 Beadle 和 Tatum 的工作。现在让我们仔细地看看他们的工作。

在生物体系统内，大体上有着许多大小象氨基酸、葡萄糖、嘌呤、嘧啶的分子，以及少数大小象蛋白质或比蛋白质更大的分子化合物。所谓这些大分子和小分子化合物，它们的合成方式和生理功能是完全不同的。虽然蛋白质或核酸本身都是通过共价键把许多较小分子连接起来的一条长链，但是，从这些小分子到合成这两类不同的大分子化合物，其方式基本上不同。因为本书大部分篇幅是研究 DNA、RNA 和蛋白质的合成方式，所以本节对此不多赘述，只作如下简要说明：在酶的作用下，每次把一个一个结构元件添加到蛋白质或核酸的增长着的分子链上，从而就合成了这些大分子化合物。然而，颇有趣味的是，接到长链上去的一个个结构元件按其排列(顺序)来说是由模板决定的。例如，DNA 分子链其结构元件即嘌呤和嘧啶的排列顺序，总是正确无误地与模板 DNA 分子，亦即亲代 DNA 的结构元件的排列顺序完全相同。以后我们将会相当详细地讲述这一过程是如何发生的，现在我们先讲一讲小分子或称结构元件的合成步骤。

氨基酸、嘌呤和嘧啶的合成是一步一步地从简单的前体分子开始到形成产生出这些作为结构单位的最终产物分子。在这一过

程中，每一步均要通过不同酶的催化，而且不同的代谢途径其代谢步骤甚至可高达 30 个至 50 个之多。酶是蛋白质分子，这种分子在细胞中起催化作用，每一个酶都是按"遗传上的信息产生的"，它们催化一个或几个特定的反应，催化的功能是促进两个相关分子（底物与产物——译注）间的代谢反应达到平衡。在缺少催化剂的情况下，这种平衡的速率非常之慢。图 29 表示单个代谢途径或一系列相互转换的代谢途径达到恒态性的条件。当在代谢途径的一头，提供作用物的分子时，通过各个酶的催化作用，将会使得各个中间产物分子（各个分子均由各自特异性的酶来催化）的浓度逐步达到平衡。可是，由于在代谢途径中，每个中间产物都是由位于它前面的那个酶所催化而产生的，所以对一个具体的反应步骤来说，作用物与生成物之间不存在可逆性的平衡（双向平衡——译注），所能够达到的仅不过是恒态性的平衡（单向平衡——译注），亦即代谢途径的总体平衡（图 29）。

这些代谢途径的生化分析需要借助于一种或几种技术。这些技术可使单独研究各个代谢途径的每个步骤成为可能，或者有助

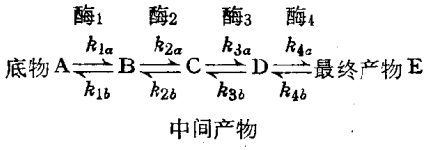


图 29　代谢途径的恒态性概念

代谢途径中的每一步中间产物，其浓度的变化将取决于各个步骤的合成反应和分解反应的速率。因为在任何一步代谢途径中，其平衡常数"实际上"恒定，所以改变任何一个化合物的浓度将会导致代谢途径中其他中间产物的浓度发生变化，在代谢途径中如果起始合成反应停止，此时所有中间产物的浓度将达到平衡状态，那末通过每一步反应的已知平衡常数就可以估算处在这一平衡状态中所有中间产物的浓度。反之，在上述情况下，如果知道每一中间产物的量，那么就可推算出所有反应的平衡常数。假定E_1, E_2, E_3和E_4的酶浓度恒定，又如果底物A持续过量，那末每一可逆反应的反应速率常数 k_1, k_2, k_3 和 k_4，在任一指定时间内，将能规定中间产物B，C和D的产量。这一比率一般保持相对恒定（指：B、C和D的浓度保持相对恒定）。甚至当A为限量供应也是如此，但这一比率亦受最终产物E被利用或除去的速率所影响，如果速率常数 k_2 和 k_3 强烈地促使 C→B 和 C→D 的反应过程，那末C的含量决不会高。另外，如果所有反应均朝着有利于生化合成最终产物E的方向进行，那么所有中间产物的含量就低。

于研究整个代谢途径中某些步骤所发生的特殊转变。例如，使用待研究的代谢途径中每个酶的专一性抑制剂，可以特异性地阻断途径中的任何所要研究的步骤，以便鉴别受阻部位之前聚积的物质，也就是测定了这个步骤的前体物质。Beadle 和 Tatum 发明了一种不用酶抑制剂并能完成同样的工作方法，他们使用遗传突变的方法得到了 *Neurospora Crassa* 许多酶功能突变型菌株，如前所述野生型 *Neurospora Crassa* 只要供给生物素就能自养。

为了说明 Beadle 和 Tatume 是如何建立他们的实验设计的，我们先从作者分离链孢霉突变品系和对其特征鉴定有预见性的实验开始介绍，这种突变品系不同于野生型，即除非在培养基中提供精氨酸的情况下，它才能生长(图 30)，当最初分离出精氨酸依赖型的突变型的时候，必须对它们进行反复的生长试验，从而鉴定它们对精氨酸的特殊需要。方法是先用 X射线处理细胞，把这些细胞接种到通常所说的完全培养基上（完全培养基含有全部氨基酸、维生素、嘌呤等)，许多营养突变型在这种培养基上均能生

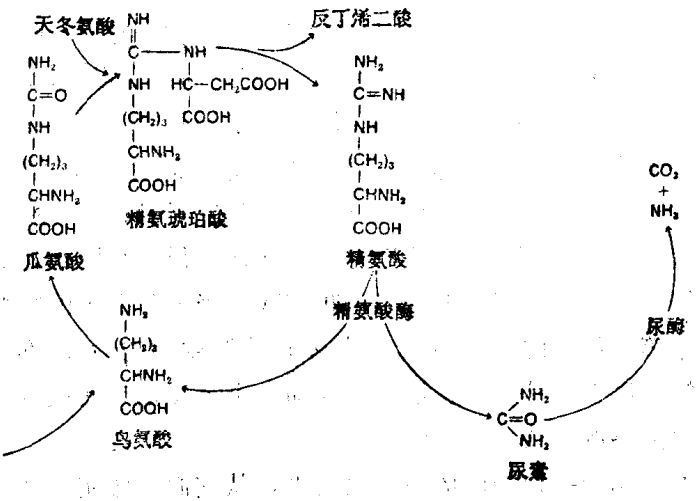


图 30　链孢霉的精氨酸依赖型

在一系列酶的催化作用下，合成了精氨酸。每一合成的步骤由于催化它的酶的失活而被阻断。酶失活的原因很多，遗传突变是其中之一，任何一步的阻断就会导致以下两种结果：(a)阻断前的前体物的积累，(b)它的生长需要精氨酸。

长，然后把它们分别接种到含有精氨酸和精氨酸的每个前体物的培养基上，测定每个突变系是否能够生长。

在分离出一个或更多的精氨酸营养突变型之后，下一步是要鉴定出在精氨酸代谢途径中的阻断位点。在还没有很好了解清楚中间代谢过程之前，如何选择最可能是精氨酸的前体的一些分子做试验不很明确，所以就应当在所有可疑的中间产物上测定每一突变型的生长特性，看它是否能生长，从而确定阻断位点。但是，这一方法未必总是可信的，因为许多细胞透不过一定的小分子，例如精氨酸前体物，精氨琥珀酸这样的小分子化合物，通常透不过链孢霉细胞(图 30)。在某些情况中，通过在代谢途径的某处测定出中间产物被累积起来，就证明为该处即是阻断的位点。这一方法也不可靠，因为积累的中间产物可以作为别的代谢途径的作用物或被降解。

为了说明如何利用突变型来研究代谢途径的中间步骤，就需要将上述试验作进一步展开，其一，需要增加代谢途径中每一个酶产生变异的机会，在这一方面现已分离出许多有关的精氨酸营养突变型。其二，每个突变型与所有另外一些突变型进行成对杂交。由同一基因突变所产生的突变品系，它们之间杂交以后，不产生野生型的后代(除了很低的频率之外)。而不同基因突变系之间杂交将会产生野生型的后代(图 31)。经突变型之间大量杂交以后，可把许多突变型分成许多等位群(一个基因之内的所有突变，相互之间是等位的，换言之，等位群数与控制代谢途径的基因数相等)。在这实例中，发现了等位群Ⅰ、Ⅱ、Ⅲ和Ⅳ。当等位群Ⅰ的突变型与另一个等位群Ⅰ的突变型杂交时，不产生野生型的后代，但是，当所有等位群Ⅰ与其他等位群中的任何一个突变型杂交便有可能产生出野生型的后代(图 31)。

通常同一等位群内的所有突变型，从它们在假定的精氨酸前体物上的生长样式来看，一般都有着相似的表现型(图31)。相似表现型的出现说明单个基因的一切突变导致改变同一个酶。由同一基因突变而来的所有突变型将有着相同的代谢缺陷和表示要求同

(a) 等位性　　代谢途径

$$A \xrightarrow{\text{酶 1}} B \xrightarrow{\text{酶 2}} C \xrightarrow{\text{酶 3}} D$$

鸟氨酸　瓜氨酸　精氨酸

(b) 等位群与生长特征

等位群 I　等位群II　等位群III与IV

$$A \xrightarrow{\text{酶 1}} B \xrightarrow{\text{酶 2}} C \xrightarrow{\text{酶 3}} D$$

鸟氨酸　瓜氨酸　精氨酸

等位群	生长情况			
	A	B	C	D
I	0	+	+	+
II	0	0	+	+
III	0	0	0	+
IV	0	0	0	+

(c) 是三种酶四个基因吗?

提出 $C \xrightarrow[\text{一个代谢步骤}]{\text{两个基因}} D$ 是经由发现催化该反应的第二个酶之后才证实的:

等位群III　等位群IV

$$C \xrightarrow{\text{酶 3}} X \xrightarrow{\text{酶 4}} D$$

瓜氨酸　精氨琥珀酸　精氨酸

图 31　精氨酸依赖型的等位群

(a) 表示产生合成精氨酸代谢途径的早期模式。A 是鸟氨酸的前体物，鸟氨酸是瓜氨酸的前体物，瓜氨酸是精氨酸的前体物，一个反应由一种酶催化。

(b) 表示等位群 I 所有的突变体的特征是要求相同的生长条件（有 A 物质不能生长，给予 B、C 和 D 物质能正常生长），并且通过实验发现它们事实上是等位的（也就是说，等位群 I 的突变型之间相互杂交，不产生重组型的后代）。然而本例中二个突变型等位群III和IV表示出要求同一个生长条件（在有 A、B 和 C 物质时不能生长，给予 D 物质后才能正常生长）。

(c) 后来发现从 C 转变到 D 需要两步过程，第一步，由瓜氨酸转变成精氨基琥珀酸，第二步是由精氨琥珀酸转变成精氨酸。每一步均通过特定酶的催化作用才得以实现。第一步是由等位群III控制，第二步是由等位群IV控制。——在代谢途径中一个基因对一个酶。

一个生长条件。在图 31 中，4 个等位突变群，从等位群 I 到等位群Ⅳ，对它们就假设的代谢途径作了缺陷位置的测定，A、B 和 C 表示最终代谢产物的前体 (A→B→C→D)。通过遗传测验，表明等位群Ⅲ和Ⅳ是不同的，但是通过生长测验，他们又是相同的，这种特殊现象在实验过程中是屡见不鲜的，然而通过进一步的观察，往往会发现在代谢途径中又增加了一个中间产物和一个酶。对于上述实例，在图 31 中的(c)有精确而简明的说明。

经过这一通用方法鉴定大量的代谢途径之后，显而易见每一代谢途径的每一步骤是由特定的酶来催化的，而且每一种酶的催化特异性都是由特定的基因所决定，这一概念就是人们熟知的一个基因一个酶的假说。在狭意上说，这一假说表示基因突变的结果导致酶的突变。广义上说，表示通过基因中包含的信息决定着酶的特异性。然而，早期的研究不可能检定出突变的真正原因，即不可能在基因水平上，或是在酶的水平上进行研究。

在对 Beadle-Tatum 的工作介绍结束之前，指出他们用链孢霉为材料所进行的另一类性质的实验是很有意义的。如果在基本培养基上 (表 1)，接种二个在遗传上不同的突变型的大量分生孢子，那末在短期内可以发现菌丝迅速生长(图 32)，这个现象可以解释为分生孢子细胞开始在培养基内外发芽，在基本培养基上每个分生孢子可产生出极小的薄壁的菌丝。如果两个不同突变系的

表 1　Vogel-Bonner 基本培养基

(适用于大肠杆菌、沙门氏菌、枯草杆菌等)

	克/升H_2O
$MgSO_4 \cdot 7H_2O$	0.2
柠檬酸	2.0
无水 K_2HPO_4	10.0
$NaNH_4HPO_4 \cdot 4H_2O$	3.5
葡萄糖(或其他碳源)	5.0

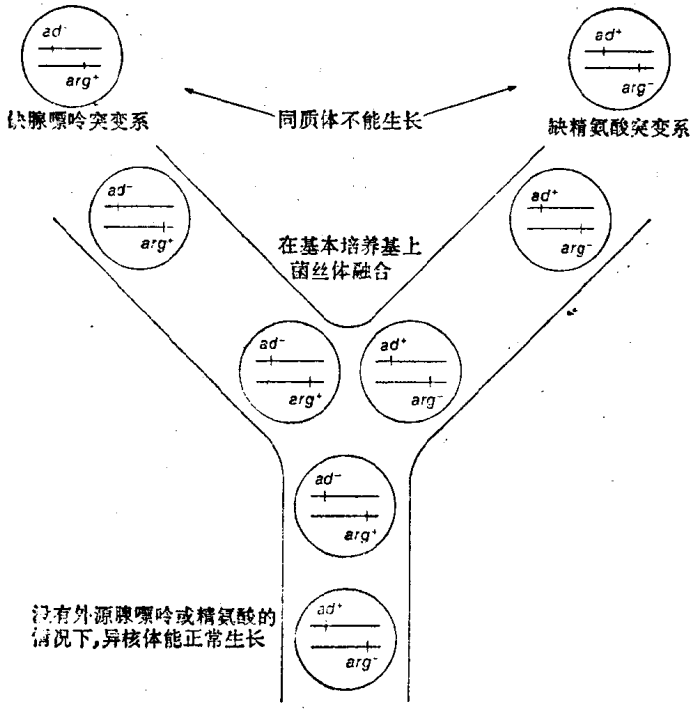


图 32 链孢霉异核体

一个异核体是指一个细胞或一个菌丝体含有两个遗传上不同类型的核。在本例中，一个核是缺精氨酸突变系 arg^-，另一个是缺腺嘌呤突变系 ad^-(对于它的生长，需要腺嘌呤)，这两个突变系中的任何一个都不能在基本培养基上生长，但是异核体都能在这种培养基上生长正常，这一事实说明在缺精氨酸突变系的核中含有腺嘌呤的基因，缺腺嘌呤的核中含有精氨酸的基因，当这样的两个核存在于同一个细胞质中，那末就全部满足了他们的生长需要。

菌丝偶然碰在一起，尤其是在营养物质缺乏的条件下，两个菌丝之间形成胞质桥，并且从这一接触点形成新的菌丝细胞，新的菌丝细胞的生长是因为它们包含有二个突变型的核，一个突变型的核在营养上能补充另一个突变型的核，反之也一样。换言之，两个突变型相互补充。这如同两辆旧的福特牌汽车，将其中一辆好的发电机装配到另一辆有着好的汽缸的汽车里一样。

当链孢霉菌丝细胞具有二个或二个以上不同遗传型的核，我们称这种菌丝为异核体，它在一定程度上类似于二倍体有机体中

的双倍性杂合子(图 33)。所不同的是，在异核体中，非等位突变基因存在于不同的核中，在二倍体中存在于同一个核中。然而两者都是由一个野生型的等位基因补充一个突变型的等位基因，从而在表型上显示野生型的特征。在二倍体中通过观察表现型常常不大可能区别 *AABB* 和 *AaBb* 的遗传型，同样，在异核体中，要把野生型 *AB* 和异核体 *Ab/aB* 之间区别开来也是困难的，因为他们均表现为野生型。

异核体测验与遗传重组测验、生长互补测验相并立，被认为是

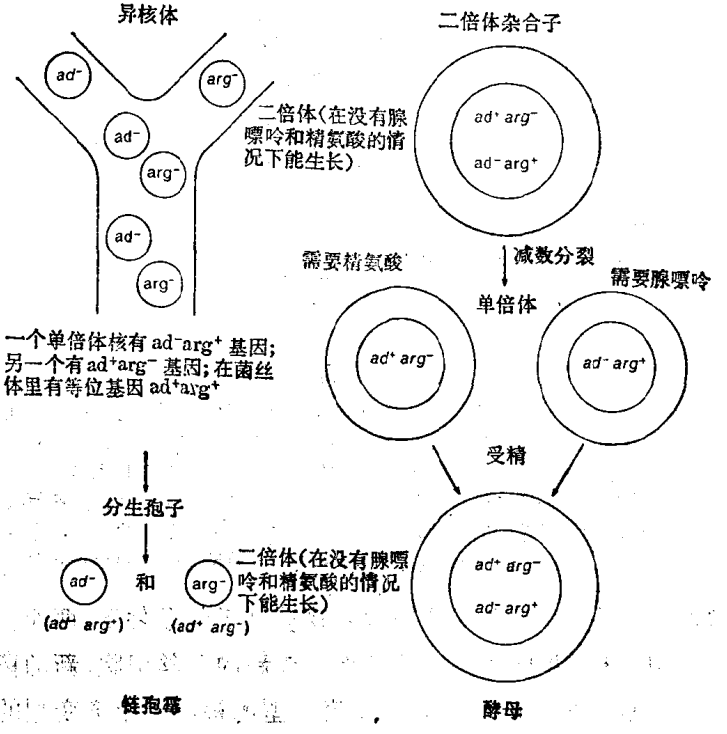


图 33 异核体和二倍体杂合子

在二倍体异质结合子中，常常用野生型或显性等位基因来标志突变型的等位基因，对异核体来说同样适用。但是，二倍体细胞中，两对等位基因是在同一个核中，然而在异核体中两对等位基因只是在不同的核中。在异质结合子的情况下，双重突变体的遗传检测，需要通过杂交测验(图 28)。但是，在异核体中，就不需要做杂交试验而能加以判别，因为由分生孢子长成的纯系菌株(同质体)只含有某一种核，因而，异核体中二种类型的核可被分开。

第三类行之有效的测验等位性的方法。事实上，若干年以前，人们把异核体测验称作为等位性功能测验，但是正如图 32 和 33 所表示的那样，通过异核体测验根本不能测定出基因功能的性质，它只能测定出二个基因的功能是相同还是不同。两个不能互补的突变型，很可能是在二个相同或十分相近的位置上所产生的突变型。二个能互补的突变型，则说明它们有着不同的缺陷。然而，如果从遗传重组的等位性测验和功能互补的等位性测验二者之间的关系来看，它们在逻辑上是互为表里的。既然遗传等位群（即互补群——译注）中的各个突变型，事实上也表现为功能上的等位群，并可以利用遗传重组测验测定出等位群中的不同突变型，那末可以设想，从理论上讲，利用功能测验也同样应该能测出等位群中的不同突变型。但是，事实并非如此，原因何在，多少年来众说纷云，争论不休。特别是后来发现基因内的遗传重组甚至与基因间的重组表现为同样频繁，从而使得对这一问题的争论达到了高潮。

为了说明在基因之内的重组概念，让我们再来看看精氨酸营养突变型的等位性例子的模式图（图 31）。等位群 I（精氨酸突变型）是"控制"酶 E_1 的特异性，此酶的作用是催化代谢途径中的第一个步骤，即催化 A→B。如果在此突变型中寻找，那么我们能分离 25 或 30 个精氨酸突变型（1, 2, 3……30）。并且通过这些突变型在有 B、C、D 但没有A物质的情况下均能生长，以及通过基因的重组测验和功能测验发现它们是等位基因。这些突变型被归入单个基因的等位群，因为在它们之中，确实没有发生重组。然而，如何来证明基因之内的重组呢？

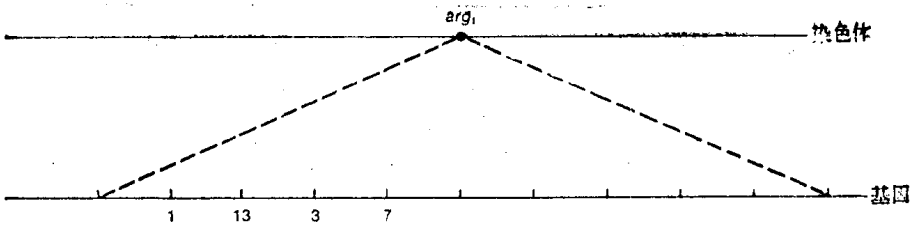
要证明基因之内的重组，我们不妨回忆一下基因之间重组频率。在一条长染色体上的任何两点之间的重组频率是随这两点之间的距离而变化。据此，凭直觉来理解，如果一个基因二端的距离非常近，那么我们可知道该基因内的重组频率很低。在实践上，当涉及测量相当短的遗传学距离的场合，必须认真对待。位于 10 个图单位距离的基因，其表现出 10%的重组，也就是说从 100 个后代群体中大约有希望出现 10 个重组类型，可是，在一个无限大的

群体中，只统计100个后代中的10个重组类型，其数目太小，可靠性低。从1000个后代中统计出100个重组类型就比较可信了。基因相距1个图距单位，就需要更大的后代数目，如果考虑到可靠性，就至少要从1000个后代中统计出10个重组类型，如能从10000个后代中测出100个重组类型，可靠程度就会更高。

在早期的大多数遗传等位性测验中，没有一个人认为在基因之内会发生交换。事实上，按摩尔根的说法，关于基因的特征之一即“基因是染色体上的一个最小的片断，在此片断内不发生交换”。如此确切无疑的定义，有着一种作用，这就是往往使人们根据此定义设计实验方案，而且当实验结果与定义不符合时，人们决不会去“怀疑”定义的正确性，而是认为自己的结果是错误的，所以当在等位突变型之间进行杂交并出现重组类型的时候，人们把它当作混杂而剔除掉，但经过仔细查看等位突变型之间的杂交，发现在后代中确实出现了极少数的重组类型。

在基因内能否产生交换的争论日趋激化，不仅是因为长期以来摩尔根表述的基因不能通过交换再分割的固有观念已经从根本上发生动摇，而且也是因为通过重组给基因下的定义和通过功能互补给基因下的定义，在相当程度上有出入。如果重组能发生在一个基因之内的一个位点上，那末有理由相信在一个基因之内的其他位点上也能发生重组。可见如果一个基因是一个单个功能单位的话，那么通过基因内的交换证明基因能再分的事实。使人确信功能单位和重组单位其实两者不是同一码事。

让我们重新来看精氨酸的突变型。如果在精氨酸(Arg_1)的等位性测验中，我们观察突变型3和13的杂交后代100个个体，发现全部后代均为突变型，与亲本类型相同，表明突变型3和13是等位的。但是，如果我们测定50万个后代，发现有25个其生长特性象野生型，那么我们将得出什么样的结论呢？第一我们应测出精氨酸缺陷型产生野生型的比例，从而可以知道25个野生型是否可能来之于回复突变($arg_1^- \rightarrow arg_1^+$)，如果排除回复突变的可能性，那么第二步就应当把30个精氨酸突变型全部两两进行配对杂交，



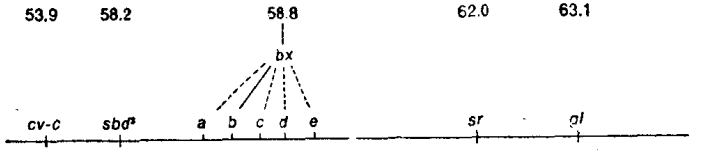
	野生型重组体 × 10^6	重组体总数 × 10^6	重组百分率	图距
$arg_1 \times arg_7$	100	200	0.0002	0.02
$arg_3 \times arg_{13}$	25	50	0.00005	0.005
$arg_{13} \times arg_7$	50	100	0.0001	0.01

图 **34** arg_1 基因遗传学图

把分离出 30 个 arg_1 的突变型，进行所有可能的成对杂交，每一个杂交组合测定约 1 百万个子代个体以检查野生型的生长情况，也就是说测定子代的重组类型，其结果每个组合的重组数都很低。但是它已证明基因呈直线排列，并且沿着一个基因长度方向内的许多位置上可以发生很多突变，此图类似于一张小型染色体图。重组体的总数是野生型重组体的两倍，因为与每个野生型对应的还有一个双重突变型(表型与单突变型相同)。

就可作出一张精氨酸的基因图（图 34)。这一基因图除了突变位点非常靠近之外，它类似于一张染色体图。

Edward B. Lewis 第一个指出在功能基因之内可以重组，因为他用果蝇作为试验材料，所以他不可能提出功能基因的确切定义，但是他通过相似的表现型可以确定功能上的等位性以及可以在染色体的很短的区段内作出许多功能基因的突变图（图 35)。*d* 和 *e* 是一个复合体"基因"，在染色体上是两个位点。如果产生的杂合体，*d* 和 *e* 在同一条染色体上，野生型等位基因＋和＋是在另一条染色体上(*de*/＋＋)，那么此种杂合体果蝇表现为野生型。如果上述等位基因重新排列，也就是在每一同源染色体上带有一个突变型等位基因(*d*＋/＋*e*)，那么就成为突变型果蝇。它似乎象 *dd* 和 *ee* 的突变型品系。当等位基因的排列是 *de*/＋＋，这叫做顺式排列，如果是 *d*＋/＋*e* 方式排列，就称反式排列。由于上述等位基因有着两种排列方式，所以使果蝇在某种场合下表现出突变型，而在另一种情况下表现出野生型的特征(图 35)。



(a)

基因内	基因间	排列方式	表现型
$\frac{de}{++}$		顺式	野生型
$\frac{d+}{+e}$		反式	突变型
	$\frac{bx\ sr}{+\ -}$	顺式	野生型
	$\frac{bx\ +}{+\ sr}$	反式	野生型

(b)

图 35　检测果蝇基因等位性的顺反子测验

(a) 此图用符号表示果蝇一条染色体上的一个小片断。图顶上部的数字表示几个基因距离染色体一端的距离，下部在染色体上的符号表示在一条染色体上的基因。Lewis 所研究的*bx*(双胸)基因在遗传上可分五个区点。每点之间相距在 0.5 到 1.0 图距单位之间。

(b) 在两种类型的杂合体中，比较顺式和反式杂合体的表现型，一是等位基因之内的突变型杂合体(*d* 和 *e*)，二是等位基因之间的突变杂合体(*bx* 和 *sr*)。前者的反式结构是突变型，后者顺式和反式均是野生型。

曾经观察了成千个顺式(*AB*/*ab*)和反式(*Ab*/*aB*)排列的遗传品系。但是 Lewis 发现凡是顺式杂合体表现为野生型，反式杂合体表现为突变型。当时对这一发现，人们都感到不可思议。实际上，这是最好地证明了 Lewis 所研究的突变是发生在功能基因之内。在顺式杂合体中，一个同源染色体携带着一个野生型基因(＋＋)，但是在反式杂合体中，两条同源染色体的同一基因都是有缺陷的(＋－和－＋)。通过表现型上相同的突变型杂交出现恢复原状的重组体后代，就证明了突变是发生在基因之内的不同位点上。

Lewis 的工作引起了许多遗传学家去研究基因之内的重组，并且基本上对研究的每一实例，只要分析大量的杂交后代都看到了基因之内发生了重组。因此，最近几年已经无人为之而操心了。这是因为功能基因在染色体上所占的范围比重组基因大，重组

基因包含在功能基因之内，因此二者不呈对应并立的关系。基因内的重组是发生在比较短的染色体片断内。讲述了 DNA 的简史以后，我们将返回到从分子的角度来阐明 DNA 的结构和功能。

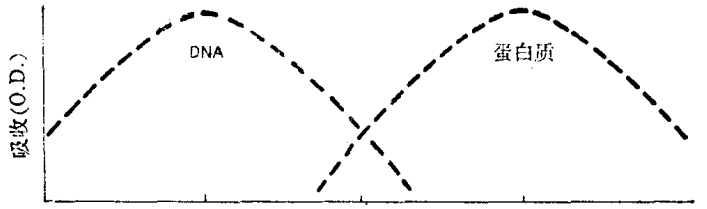
DNA 及其在遗传中的作用

追溯历史，我们看到在遗传学发展的同时，平行地存在着一系列有关核酸的发现，但当时在大多数人看来核酸与上述遗传现象毫无关系。其实核酸的发现多数是在 1944 年以前，但是后来才逐渐地把 DNA 和基因挂起钩来，并且它们之间的联系越来越清楚。对于 DNA 和基因之间的关系已家喻户晓的今天，人们不禁要问：为什么要花那么长的时间才发现 DNA 是遗传物质以及当发现后为什么遗传学家们要花那么长的时间才承认这一事实，这是不能理解的。

在一百多年前，第一次分离出了非结晶白粉末状的 DNA，直到许多年以后还不知道 DNA 和 RNA 是两种不同的核酸，而且到 1944 年人们还否认DNA是遗传物质。然而在十九世纪末二十世纪初已经知道细胞核含有丰富的核酸，后来进一步又发现核酸存在于染色体中。

到 1940 年，有二个重要工作把基因和 DNA 联系了起来，但是在当时，大多数遗传学家要么是不考虑基因的化学组成，要么推测基因是蛋白质。其中一个工作是发现紫外光能诱发基因突变。L. J. Stadler 用玉米为材料，分别用业已知道的 DNA 和蛋白质的最大吸收波长的紫外光作为诱变因素(突变诱变剂)，以检查它们的诱变效应。蛋白质吸收紫外光的高峰是 280毫微米，然而DNA 的吸收高峰为 260 毫微米，实验证明诱发突变最有效的波长为 265 毫微米左右(图 36)。上述结果虽然决没有证明 DNA 是遗传物质，但是他们之间的内在联系已是很明显了。

第二个工作发现配子具有单倍体染色体数，同一种生物，配子中 DNA 的含量恰好是体细胞的一半。同时还发现细胞中的蛋白质和 RNA 的含量与染色体的数目不是一对一的对应关系，但后



波长, nm	细胞中的突变数	存活者, %
260	50-100	60
270	1-5	95
280	0-1	100

图 36　通过紫外光诱发突变

Stadler 试验了不同紫外光波长诱发突变效应，他发现 DNA 的吸收高峰作为诱变剂和钝化剂的效应，比蛋白质的吸收高峰显著有效。

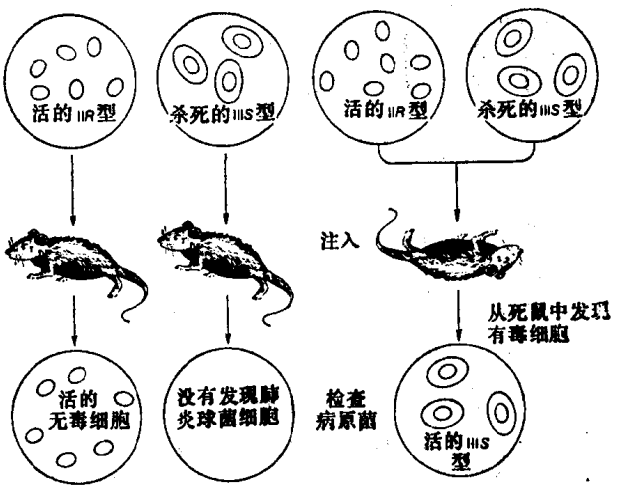
来证明了细胞染色体蛋白质，即所谓组蛋白的含量与 DNA 一样随染色体数目变化，DNA 局限于细胞核中，RNA 和蛋白质存在于整个细胞中，这一事实暗示 DNA 是遗传物质，但这一点当时没有被人们所认识(六十年代早期，发现许多细胞器如叶绿体、线粒体中存在着 DNA，但当时对 DNA 在遗传上的作用已经看清楚了)。

虽然 Beadle 和 Tatum 的工作证明了基因与代谢缺陷之间存在着一对一的对应关系，但是他们的工作丝毫没有暗示出 DNA 是遗传物质。上述二个工作虽然可被他们和其他遗传学家所利用，但当时普遍相信基因是蛋白质组分，而不是核酸，其原因是蛋白质的字母表上有 20 种"字母"(20 种氨基酸)，它可提供的遗传信息对于解释在自然界中产生变异性的原因来说已绰绰有余。然而核酸的字母表只有 4 种字母(2 个嘌呤和 2 个嘧啶)，仅能提供有限的遗传信息的可能性。

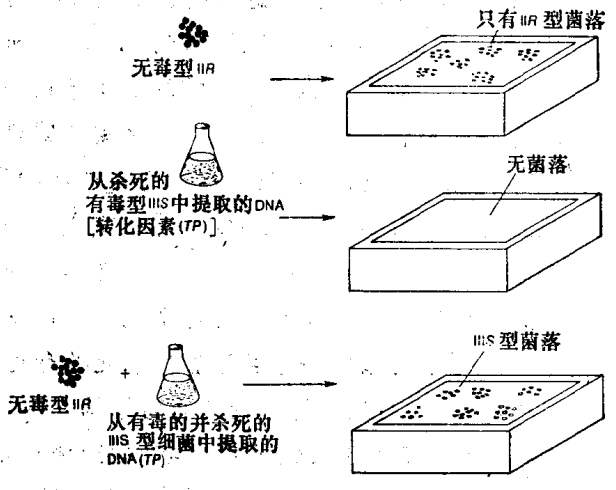
在 1928 年，英国的一个内科医生 F. Griffith，他不是一个遗传学家，因此不关心遗传学问题，但是他仍然做了遗传学方面的实验工作，并且通过他的工作第一次使人信服地证明了基因是 DNA

分子。Griffith 用肺炎球菌做实验，这种细菌是大多数哺乳动物产生肺炎的原因，他分离出不同种类的肺炎球菌品系，一种是有毒型的，一种是无毒型的。他用活的无毒型的肺炎球菌和加热杀死的有毒型肺炎球菌一起复合感染活的家兔，发现家兔感染肺炎而死亡。从有病的或死亡的家兔体内分离出活的有毒细菌细胞。无毒型的细菌或加热杀死的有毒型细胞分别单独感染家兔，决不会产生有毒的细菌细胞，只有当无毒型的细菌细胞在加热杀死的有毒细胞的参与下一起感染家兔时才能产生有毒型细胞。他从这个实验结果得出结论，无毒型的细菌细胞通过加热杀死的有毒型细菌转化成有毒型。Griffith 没有把加热杀死的有毒细菌细胞分成糖、蛋白质、DNA 等等，进行单因子转化试验，但是他推测转化的"关键"与细菌的细胞壁有关，并且认为很可能是多糖。随后的十年中，有几个细菌学家重复了 Griffith 的实验，但是遗传学家并不知道 Griffith 的工作。

O. T. Avery 和他的同事在 1944 年做了一系列的实验，从而发展了 Griffith 的实验方法。某些人认为 Avery 的实验标志



(a)



(b)

图 37 肺炎球菌的转化

(a) 表示在1928年 Griffith 实验的基本模型，他用两个肺炎球菌品系，一个是有毒的(IIIS 型)和另一个是无毒的(IIR 型)，发现小鼠经无毒细菌感染后，在其身上能找到无毒的肺炎球菌但小鼠安然无恙。用加热杀死的有毒细菌去感染小鼠，在其体内找不到任何肺炎球菌，但是如果用活的无毒细菌和加热杀死的有毒细菌一起注射入小鼠体内，小鼠就会引起发病和死亡，而且从死亡的小鼠身体内能找到有毒的细菌。(引自 Sutton, Genes, Enzymes and Heridited Disease, Rinehart and Winston, New York, 1961。)

(b) Avery 的实验模型，Avery 没有用小鼠为材料，做了同样性质的实验，重复了 Griffith 的工作并发展了实验方法。他把细菌培养在琼脂平板上。活的无毒的细菌在琼脂培养基的表面产生出 IIR 型的菌株，加热杀死的有毒细菌在琼脂培养基表面不产生菌株，但是无毒的细菌与从加热杀死的有毒细菌中提出的 DNA 混合培养产生出 IIIS 型有毒的菌株。从 IIIS 型提取的 RNA、蛋白质或碳水化合物都不能起转化的作用。

着分子生物学和分子遗传学的开始，事实上，有一个卓越的分子遗传学家在提到 Avery 时曾尊称他为分子生物学的鼻祖。Avery 和他的小组在转化无毒细胞成有毒细胞的实验中，发现活的转化物质是 DNA。这一事实，博得了人们的广泛赞赏(图 37)。

他们把有毒的肺炎球菌分离出来的各种大分子物质，一个一个地与无毒细菌培养在一起，来测定每个物质的转化作用，结果发现唯一起转化作用的是 DNA 分子。这类实验他们从两方面进

行，一种为正实验即把分离出的每一种物质直接测定其转化作用；另一种为负实验，即一次一个有选择地破坏每一种物质，然后进行转化试验，他们把某种物质的缺少与转化活性的丧失进行对比，发现凡是缺失 DNA，就伴随着出现转化作用的消失(DNA 用 DNA 酶来破坏，这是一种消化 DNA 的水解酶)。在以后的几年中，他们尽力进一步纯化 DNA 的组分并用来做转化实验，而且发展到对许多种菌类和许多遗传特征进行了转化试验。于是在相当短的时间内，几乎无人怀疑许多基因通过 DNA 都能够从一个细菌品系转移到另一个细菌品系中去。从这些实验可以推论，作为一个实体单位的基因，例如有毒基因和从有毒型分离出来的 DNA 是一样的东西。从一个细菌品系分离出的 DNA(基因)进行转化试验，如果没有被第二品系(受体)降解，那末可以被此品系摄取，一旦进入细菌细胞内，这一新进入的 DNA 能够并入到宿主的染色体上，这个并入的机制有点象真核类同源染色体之间的交换，关于这一问题，尚待以后叙述。

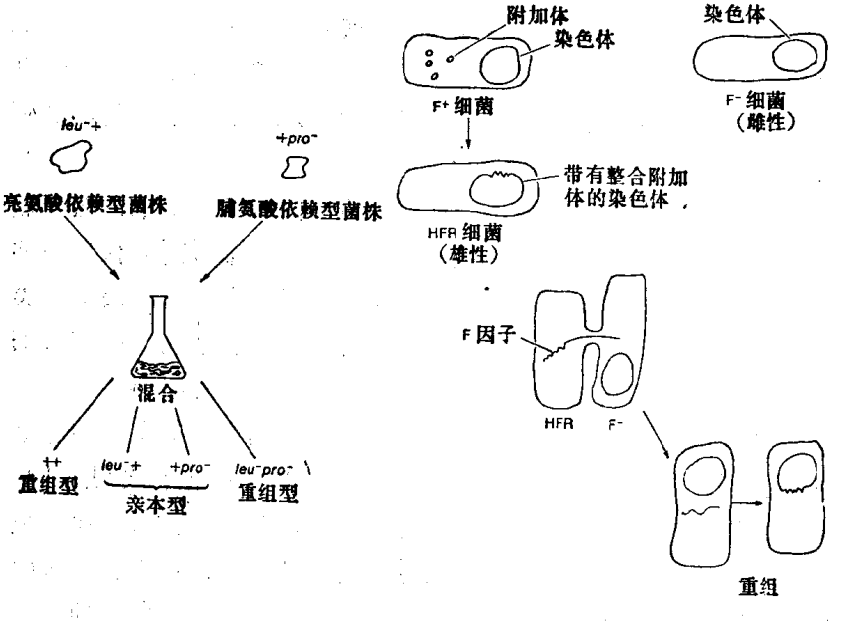
于是，到了 1945 年，DNA 作为遗传物质就有了三个重要的工作，即使如此，但是又经过了七年才使得广大遗传学家信服这一事实。在此期间，产生了新的研究思路和进入到了分子遗传学的研究水平。

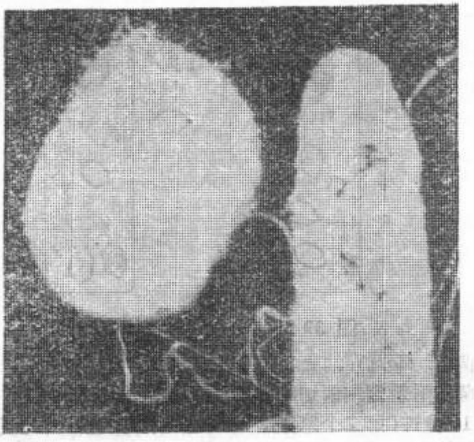
发展之一就是利用原核生物作为研究遗传学的对象，这一点在 DNA 的研究史上是有口皆碑的。在遗传学简史中，仅仅叙述了真核类生物，然而，上述关于 DNA 的研究进展，反映出原核类生物有利于进行遗传分析和极便于进行化学分析。到 1944 年，细菌，对于从事遗传学的人来说，已成为有兴趣的研究材料。在 1946 年，Tatum 和 Joshua Lederberg 证明在细菌中也能发生遗传重组，换言之，有性过程的现象同样在最简单生物中也有表现。Tatume 和 Lederberg 没有观察到细菌细胞有交配的情况，但他们根据作为遗传重组现象基础的有性过程的遗传定义，认为在细菌中也有性别。细菌的生化突变型类似于早期在链孢霉中发现的生化突变型，用两个遗传型不同的生化突变型细菌细胞混合以后，在后

代中可以分离出具有双亲基因的新类型(图 38)。

在论述证明 DNA 是遗传物质的另一个精确试验以前，有必要先探讨一下原核类遗传的另一特点，即病毒(称噬菌体或称细菌噬菌体）与所感染的细菌细胞之间的相互关系。大约在 1915 年，病毒学家和微生物学家对噬菌体已进行了研究，但是直到本世纪中叶有关噬菌体感染细菌细胞的方式或杀死细菌细胞的方式知道不多。后来，通过噬菌体颗粒作为有效的媒介把一个细菌的遗传物质转移到另一个细菌上去的所谓细菌的转导这一重要发现，才得以查明噬菌体感染细菌的方式。要知道细菌基因是如何通过噬菌体进行转移的，那么就必须要搞清楚噬菌体感染细菌的两种不同基本形式。

第一种感染类型称作为烈性感染或称溶菌性感染，图 39 可说明这一感染的大致情况，一个噬菌体颗粒首先附着到细菌的细胞壁上，于是病毒的“染色体”进入到细菌的细胞中，而病毒的外壳留





(c)

图 38　细菌的重组

(a) 两个细菌混合起来连续繁殖几代，便可发现一些后代发生了遗传重组，这是 Tatum和 Lederberg 进行的这类实验的示意图。

(b) 大体上可把细菌分为供体和受体两种，在细菌连接的过程中，供体细胞贡献出 DNA 到受体细胞中去。一般在供体细菌的群体中，有两种类型的细菌，一种称为 F^+细菌，在这种细菌细胞中有着两类 DNA，其一，大的 DNA 是呈环状染色体，其二，小的 DNA 环状分子称附加体。另一种称 HFR 型细菌细胞(高频转导菌株)，它至少带有一个整合到大环形染色体上的附加体 DNA 分子，这种类型的细菌细胞在两个细胞连接的过程中，作为"雄性"供体。

(c) 表示两个细菌细胞接合的照片。在接合过程中，HFR 菌系的环状染色体常常在整合的附加体基因(F 因子)末端处断裂(附加体最后进入受体)成线状 DNA，这一过程的发生大约(当 HFR 的整个染色体进入到受体细胞中)需要 90 分钟。整个染色体的转移是十分罕见的，仅仅是染色体的片断转移，而转移的染色体片断的长短则又取决于结合过程中周围环境阻碍接合的因素。然而，一旦染色体片断进入到 F^- 细胞，那么它就可以整合到 F^- 染色体上 (其机制相当于染色体交换)，结果在供体和受体基因之间产生重组。(由 E. L. Wollman 提供。)

在外面。在细菌细胞里的病毒染色体开始产生新的病毒颗粒，而同时使细菌本身的繁殖活动停止，使细菌的代谢改变，以及使得细菌的染色体解体。大约感染 30 分钟后，细菌细胞被溶解，也就是说，细菌细胞破裂，并释放出几百个病毒后代。

第二种感染类型，称为温和性感染。图 40 表示这种感染的大概情况。在温和性感染类型中，当噬菌体的染色体即将进入到细

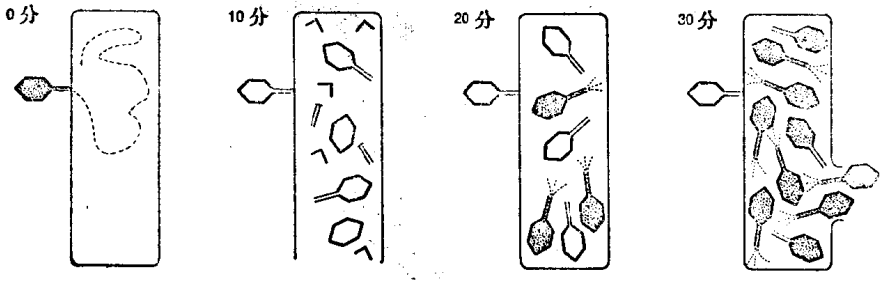


图 39 烈性溶菌噬菌体感染

感染大肠杆菌(*E. Coli*) 细胞的 T_2 噬菌体的生活周期。T_2 噬菌体的第一个感染期是噬菌体附着到细菌细胞的表面。噬菌体的染色体进入细菌细胞中，而后大约感染 10 分钟(第二个感染时期)，噬菌体染色体开始复制以及可以观察到少数未成熟的新的噬菌体。第三个时期，大约感染后 20 分钟，产生出感染性的噬菌体颗粒(具有蛋白质的外壳，内含染色体)。大约感染后 30 分钟为第四时期，此时释放出大量子代噬菌体颗粒。(引自 Stent, The Multiplication of Bacterial Viruses, Scientific American, May 1953。)

菌细胞时，其感染方式与溶菌性感染相同。但是当噬菌体的染色体进入细胞后，通过类似于交换的机制整合到细菌的染色体上，而且可以在细菌中保持许多世代。这种整合的噬菌体染色体称为原噬菌体，这种原噬菌体通常在任何时候用某种方式刺激以后，如用紫外线照射，就可以使它从细菌染色体上诱导解脱出来。原噬菌体从细菌染色体上释放出来以后，即开始与上述溶菌性感染相同的程序行事。也就是说，形成新的噬菌体颗粒，宿主细胞裂解并释放出许多成熟的噬菌体颗粒。

Lederberg 和 Norton Zinder 把上述两种感染方式结合起来研究，他们发现某一遗传型的细菌细胞裂解以后，释放出来的噬菌体能够把裂解细胞中的基因转移到一个不同遗传型的细菌细胞中，这类原属烈性感染类型的噬菌体在这种细菌细胞中就变成温和性感染类型(图 41)。这一结果说明，当在供体细胞内形成噬菌体颗粒的时候，宿主染色体的小片断被并入到这些噬菌体的颗粒内。在感染受体细胞时，细菌的染色体小片断同噬菌体染色体一起进入到受体细胞中。但是，因为在第二宿主中，它便作为原噬菌体而表现温和性类型，所以宿主的染色体不但不会被破坏，而且由

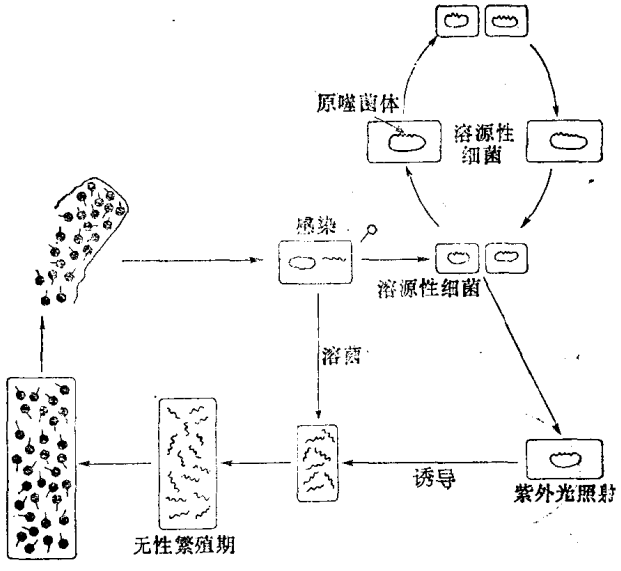


图 40　温和噬菌体感染

温和噬菌体的生活周期除了温和噬菌体的染色体整合到细菌的染色体上之外，与烈性噬菌体的生活周期相同(图 39)，换言之，温和性噬菌体的染色体其作用如同一个附加体，即类似于 F 因子(图 38)。通过紫外线照射后，病毒染色体或称原噬菌体能从细菌染色体上解脱，并且一旦游离出来，就会象烈性噬菌体一样繁殖出许多新的噬菌体。(引自 Lwoff, Bact Rev, 17:269, 1953。)

噬菌体带进来的供体细胞的某些染色体片断也能整合到受体细胞的染色体上。

与 Lederberg 和 Zinder 发现转导同年，另有一些实验证明，通过噬菌体注入到细菌中的物质是 DNA。Alfred Hershey 和 Martha Chase 于 1952 年曾成功地做了这些实验；他们在放射性^{32}P 和^{35}S 存在的情况下，让噬菌体进行繁殖，并把这种具有充分放射性物质的噬菌体去感染无放射性的细菌，当大多数噬菌体的染色体注入到细菌细胞之后不久，此时恰好是感染数分钟之后，把细菌放在组织捣碎器上充分摇动可以使噬菌体颗粒的外壳从细胞壁上分离下来，然后化验分析空的噬菌体颗粒(外壳)和细菌细胞的放射性物质，结果发现全部^{32}P 存在于细菌细胞中，^{35}S 存在于噬菌体的外壳中(图 42)。

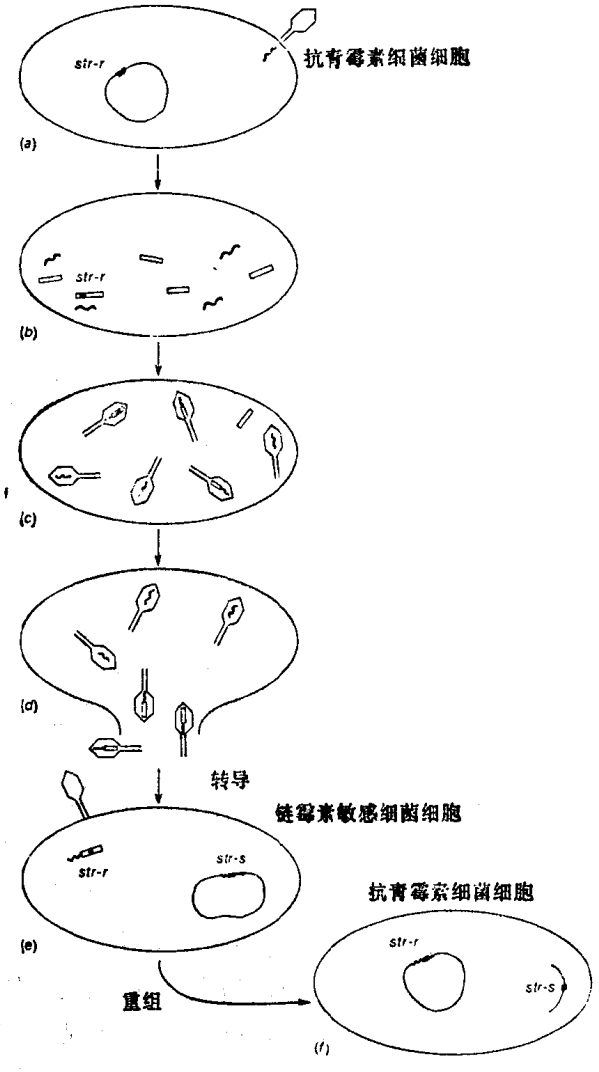


图 41 转导

此图表示从一个细菌的 DNA 转移到另一个细菌之前和之后的转导过程。(a) 通过噬菌体感染抗链霉素细菌，产生出转导噬菌体。(b) 细菌 DNA 裂解和噬菌体 DNA 复制。(c) 形成成熟的噬菌体颗粒，其中有一些噬菌体携带有宿主的某些 DNA 片断。 (d) 噬菌体大量繁殖，使供体细菌裂解。(e) 释放出来的噬菌体作为溶源性噬菌体颗粒进入到对链霉素敏感的受体细胞中。(f) 在少数重组细胞中，*str-r*（抗链霉素）的基因确实并合到宿主基因组上。(引自 Sager 与 Ryan, 1961。)

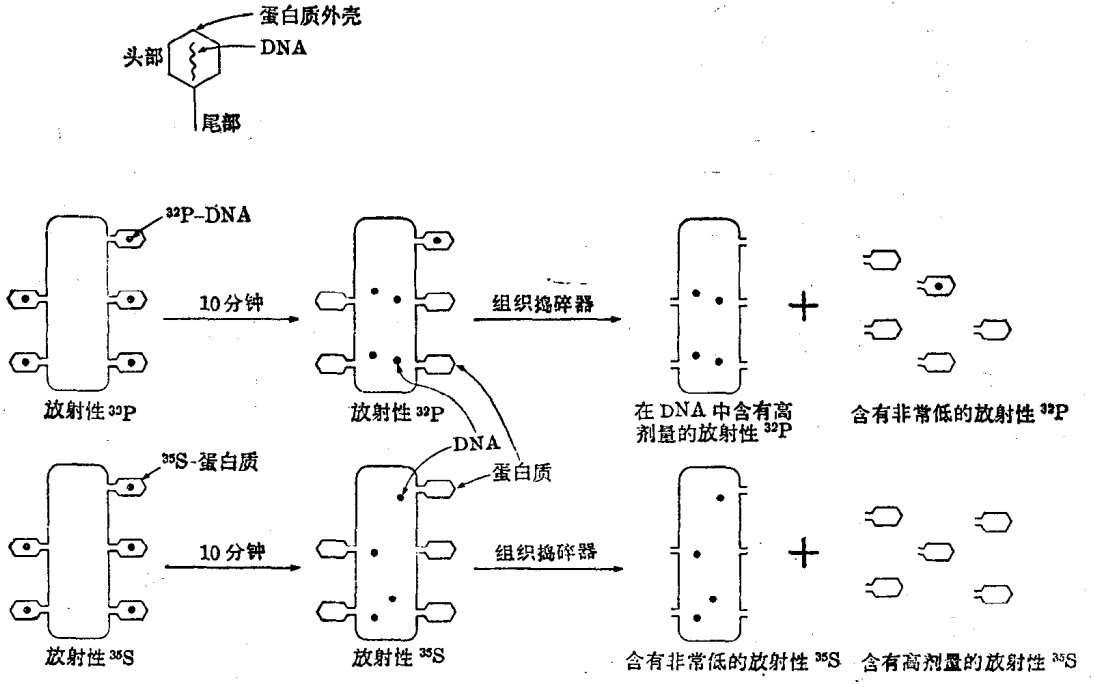


图 42　噬菌体染色体是 DNA

此图表示 Hershey 和 Chase证明噬菌体染色体是 DNA 的实验概况。用^{35}S标记蛋白质中的半胱氨酸和甲硫氨酸，^{32}P标记DNA中的磷酸(所有DNA的主链部分)。而后用含有^{32}P 和^{35}S 的噬菌体去感染细胞，并在 Waring blender 中摇动它们，为的是把"空的"噬菌体切下来，发现含^{32}P 的 DNA 在细菌中，^{35}S在噬菌体的空壳中。

因为^{32}P是结合在DNA主链的磷酸骨架上，^{35}S(作为含硫氨基酸的形式)结合在蛋白质上，所以Hershey和Chase断定噬菌体的DNA注入了细菌细胞中，而噬菌体的蛋白质却留在细菌的外面，换言之，噬菌体的染色体是DNA，噬菌体的外壳是蛋白质。图43表示一个成熟的T_4噬菌体颗粒，概括起来，这个实验告诉我们病毒的DNA携带着遗传的指令，这不仅仅使宿主的染色体解体，而且也是合成新的噬菌体颗粒的基础。在这个实验之后，就再也没有任何遗传学家怀疑DNA在遗传上的重要意义了。

在Lederberg、Zinder、Hershey和Chase实验的前三年，生物化学家Erwin Chargaff发表了关于论述DNA化学组成的论文，在那段时间里，发展了从不同种类的细胞中分离并纯化

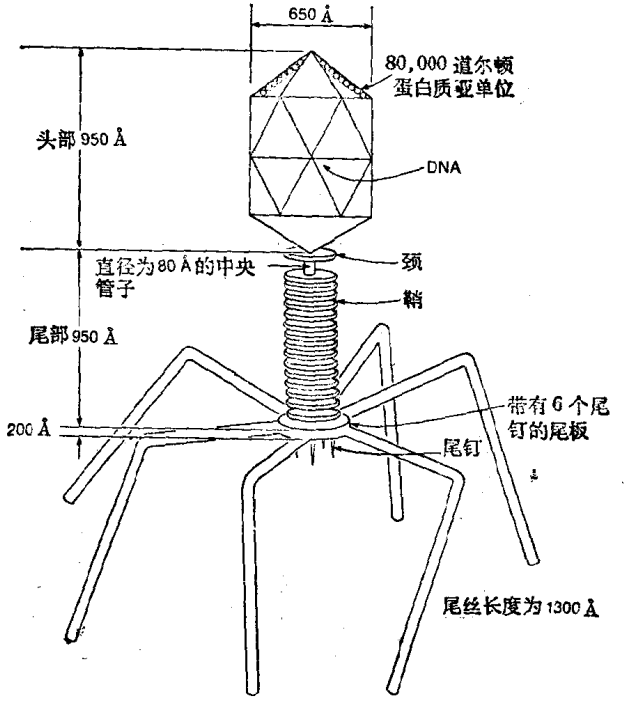


图 43 一个成熟的 T_4 噬菌体颗粒

在此例中所举的T_4噬菌体颗粒是一个比较复杂的病毒，到目前为止还不知道在噬菌体中的DNA染色体是如何包装在头部的。

DNA 的多种技术，图 44 即是这些技术之一。到四十年代末，人们才知道所有 DNA 分子主要是由脱氧核糖和磷酸组成，并发现 DNA 分子之间的差别在于其两个嘌呤和两个嘧啶的相对量不同。但是 Chargaff 发现上述碱基相对量的不同是因为生物的种类不同。也就是说，所有 DNA 不管它的来源如何，其必然有两种嘌呤，即腺嘌呤(A)和鸟嘌呤(G)，以及两种嘧啶，即胸腺嘧啶(T)和胞嘧啶(C)，鸟嘌呤的克分子数总是和胞嘧啶的克分子数相等，腺嘌呤的克分子数总是与胸腺嘧啶的克分子数相等。在所有 DNA 分子内，A＋C＝G＋T，换言之，不同种的 DNA 的差别在于 GC 和 AT 碱基对的比值不等(表 2)。1952 年以后，这一事实成了十分重要的历史大事。

1953 年 James Watson 和 Francis Crick 提出了关于 DNA 分子结构的正确模型(图 45)。DNA 分子的全部构型为由二条半分子单链组合而成的螺旋双链，单链之间彼此以碱基对之间形成的氢键相连。DNA 模型表明：腺嘌呤总是与胸腺嘧啶配对，鸟嘌呤总是与尿嘧啶配对。

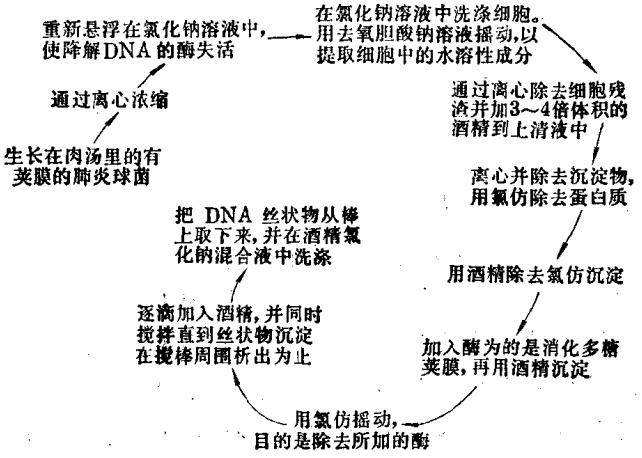


图 44　从肺炎球菌中提取 DNA

O. T. Avery, C. M. Macleod 和 M. McCarty 早期进行转化实验是按上图顺序提取 DNA，此法提取的 DNA 其杂质低于 1%。现在对于上述提取 DNA 的原型作了多种多样的改良。

表 2　不同种属生物体 DNA 分子中的 AT/GC 比例

DNA 来源	$\frac{A+T}{G+C}$
T_2 噬菌体	1.84
λ 噬菌体	1.06
大肠杆菌(*E. coli*)	0.97
芽孢杆菌(*Bacillus megaterium*)	1.66
酵　母(*Saccharomyces cerevisiae*)	1.80
黑　霉(*Aspergillus niger*)	1.00
绿　藻(*Scendesmus quadricavda*)	0.57
硅　藻(*Rhabdonema adriaticum*)	1.71
小　麦	1.22
海　胆(*Paracentrotus lividos*)	1.86
飞　蝗(*Locusta migratoria*)	1.41
鳟　鱼	1.34
家　鸡	1.36
马	1.33
人	1.40

对于 Watson 与 Crick 所提出的上述 DNA 模型以及该模型在生物学中的重大意义，今天在生物学界中仍持怀疑的人已是几乎没有了。双螺旋模型经受住了时间的考验，在这一认识的指导推动下，人们开展了数千次有益的实验研究，从而进一步澄清了遗传信息据以编码、贮存、转录以及转译的工作方式。迄今为止，人们对双螺旋结构模式的原型，虽有稍微修正，但仅仅属于旁支末节之类，模型的中心主题则仍然保持不动。所作的修正有以下三类：(1)并非所有的 DNA 分子都一无毫外地是双螺旋型式，某些微型噬菌体的 DNA 染色体分子为单链 DNA，因此亦不显示A+C=G+T 的比例关系；(2)鸟嘌呤与胞嘧啶碱基对之间有三条氢键而不是在双螺旋结构的最初模型中所讲的只有二条氢键；(3)从

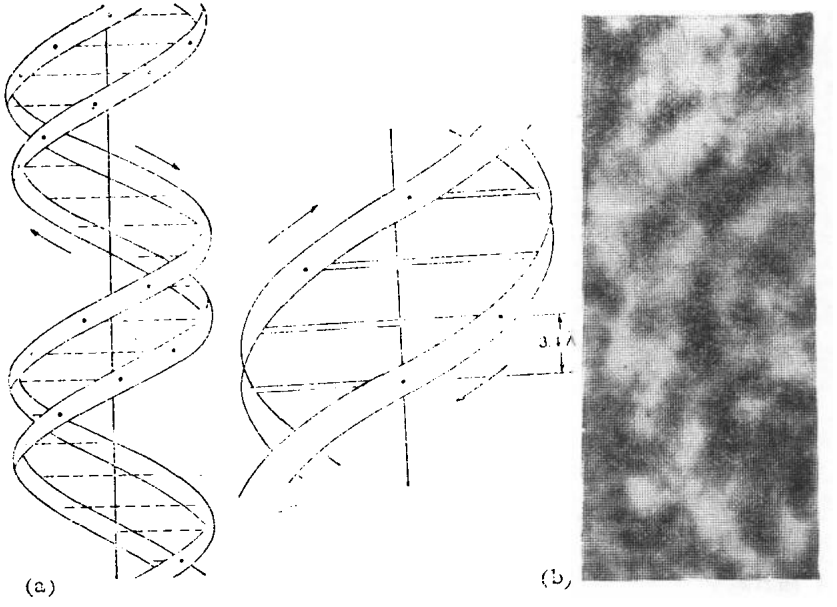
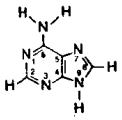
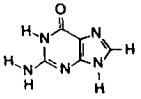
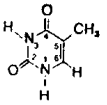
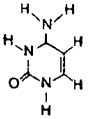
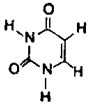
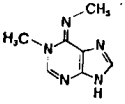
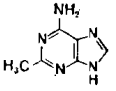
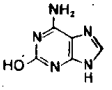
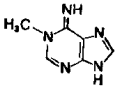


图 45 DNA 的一般结构

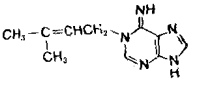
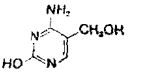
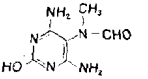
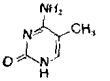
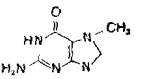
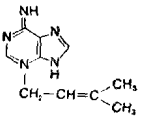
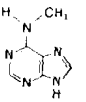
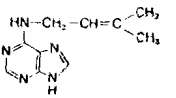
(a) DNA 是由二条单链组成的双螺旋结构的大分子化合物，单链结构中脱氧核糖分子与磷酸分子交互排列。两条单链又通过嘌呤与嘧啶之间的氢键使彼此对偶相连。图中二条单链之间的横线表示 A-T 和 C-G 碱基对之间的氢键，单链（条带）表示脱氧核糖-磷酸骨架。 (b) 图示用高分辨力电镜直接拍摄的 DNA 双螺旋链可观察的最佳图象。此图片虽然说不上是补充证据，但却引人联想，这或许可成为最终能直接观察双螺旋结构的先声。（由史坦福大学生化系 Jack D. Griffith 赠送。）

五花八门的生物种属来看，DNA 分子中的碱基成分，除 A、T、G、C 之外，偶尔也会包含另一些稀有碱基(图 46)。例如在小麦的胚和细菌中发现有 5-甲基胞嘧啶存在，在某些细菌和噬菌体中则包含有 6-甲基氨基嘌呤。此外尿嘧啶和 5-羟甲基胞嘧啶在感染大肠杆菌的偶数 T 噬菌体(如 T_2 和 T_4)中亦有发现。

作为 DNA 分子结构的全部嘌呤嘧啶和脱氧核糖，都是在前面所说的这类具有中间代谢特点的各种代谢途径中进行合成的。虽然代谢途径都受为数甚多的基因集体指导并由多种多样有关的酶给予催化完成(图 47)。

化　　合　　物	缩写名称	结　构　式
普通碱基		
腺嘌呤	A	
鸟嘌呤	G	
胸腺嘧啶	T	
胞嘧啶	C	
稀有碱基		
尿嘧啶	U	
1-甲基-N^6-甲基腺嘌呤	1,6Me_2Ade	
2-甲基腺嘌呤	2MeAde	
2-羟基腺嘌呤(异鸟嘌呤)	2(OH)Ade iso Gua	
1-甲基腺嘌呤	1MeAde	

（续上）

化　　合　　物	缩写名称	结　构　式
1-(Δ²-异戊烯基)腺嘌呤 〔1-(γ,γ-二甲基烯丙基)腺嘌呤〕	1iPeAde	
5-羟甲基胞嘧啶	5HmCyt	
2,6-二氨基-4-羟基-5-N-甲基甲酰胺基嘧啶 (6-氨基-5-N-甲基甲酰胺基异胞嘧啶)		
5-甲基胞嘧啶	5MeCyt	
3-甲基胞嘧啶	3MeCyt	
7-甲基鸟嘌呤	7MeGua	
3-(Δ²-异戊烯基)腺嘌呤 (三刺皂荚碱)	3iPeAde	
N⁶-甲基腺嘌呤(6-甲基氨基嘌呤)	6MeAde	
N⁶-(Δ²-异戊稀基)腺嘌呤 [N⁶-(γ,γ-二甲基烯丙基)腺嘌呤, 6-(3-甲基-2-丁烯基氨(基)嘌呤]	6iPeAde	

（续上）

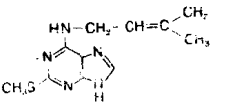
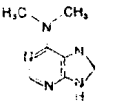
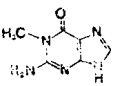
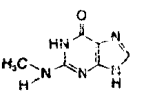
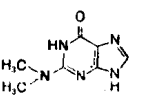
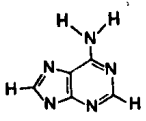
化　　合　　物	缩写名称	结　构　式
N^6-(Δ^2-异戊烯基)-2-甲基硫代腺嘌呤	6iPe2MeSAde	
N^6, N^6-二甲基腺嘌呤 (6-二甲基氨基嘌呤)	6Me2Ade	
7-甲基腺嘌呤	7MeAde	
1-甲基鸟嘌呤	1MeGua	
N^2-甲基鸟嘌呤 (6-羟-2-甲基氨基嘌呤)	2MeGua	
N^2, N^2-二甲基鸟嘌呤 (α-二甲基氨基-6-羟基嘌呤)	$2Me_2Gua$	

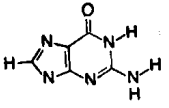
图 46　DNA 分子中的一般碱基和稀有碱基

在不同种属生物体的DNA中，最常见的四种碱基是腺嘌呤、胸腺嘧啶、鸟嘌呤和胞嘧啶。图中所列的稀有碱基，在一定种属的生物体中能够置换一种或多种普通碱基。

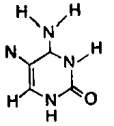
碱基



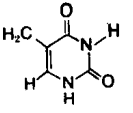
腺嘌呤



鸟嘌呤

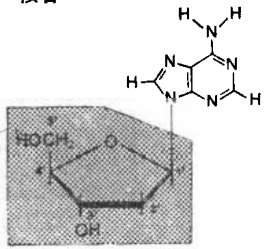


胞嘧啶

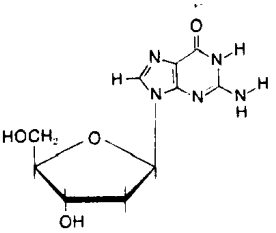


胸腺嘧啶

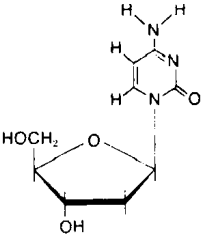
核苷



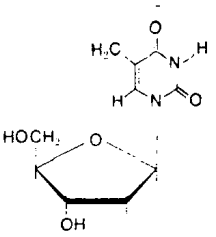
脱氧腺苷



脱氧鸟苷



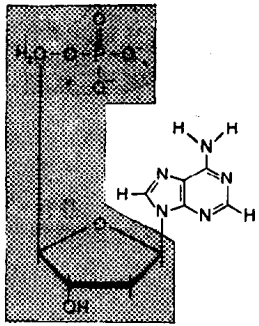
脱氧胞苷



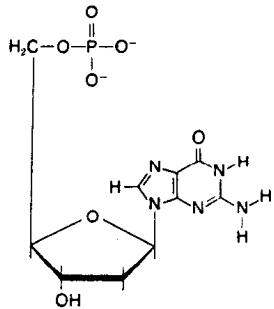
脱氧胸苷

(a) (1)

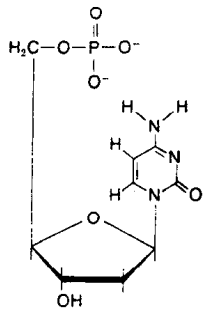
核苷酸



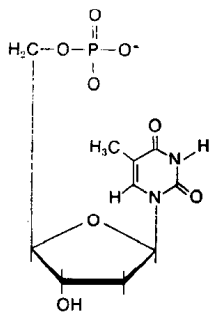
脱氧腺苷单磷酸
dAMP
(脱氧腺苷酸)



脱氧鸟苷单磷酸
dGMP
(脱氧鸟苷酸)



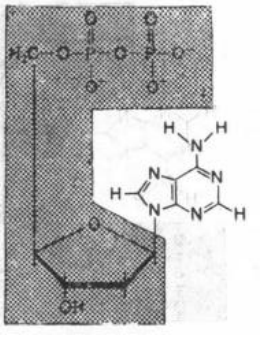
脱氧胞苷单磷酸
dCMP
(脱氧胞苷酸)



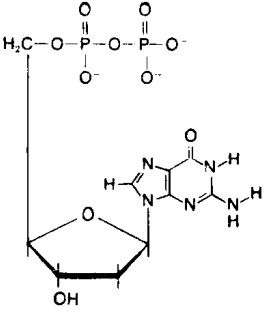
脱氧胸苷单磷酸
dTMP
(脱氧胸苷酸)

(a)(2)

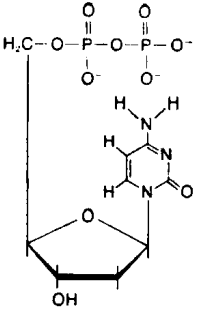
二磷酸核苷酸



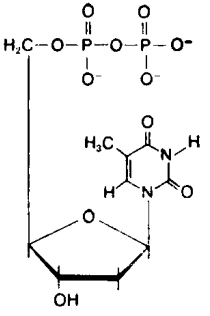
脱氧腺苷二磷酸
dADP



脱氧鸟苷二磷酸
dGDP

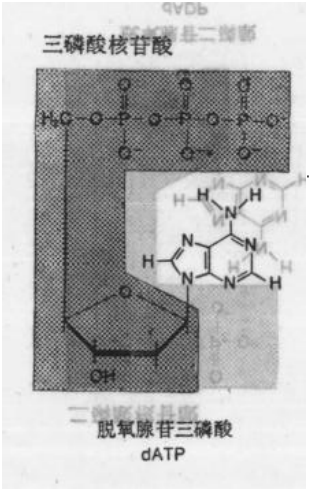


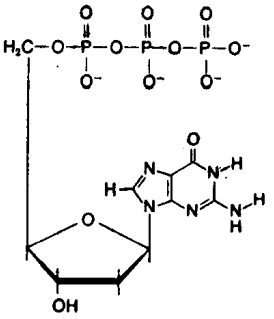
脱氧胞苷二磷酸
dCDP

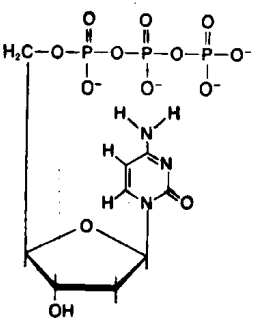


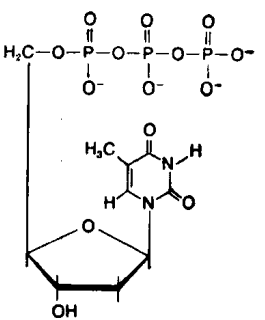
脱氧腺苷二磷酸
dTDP

(a) (3)

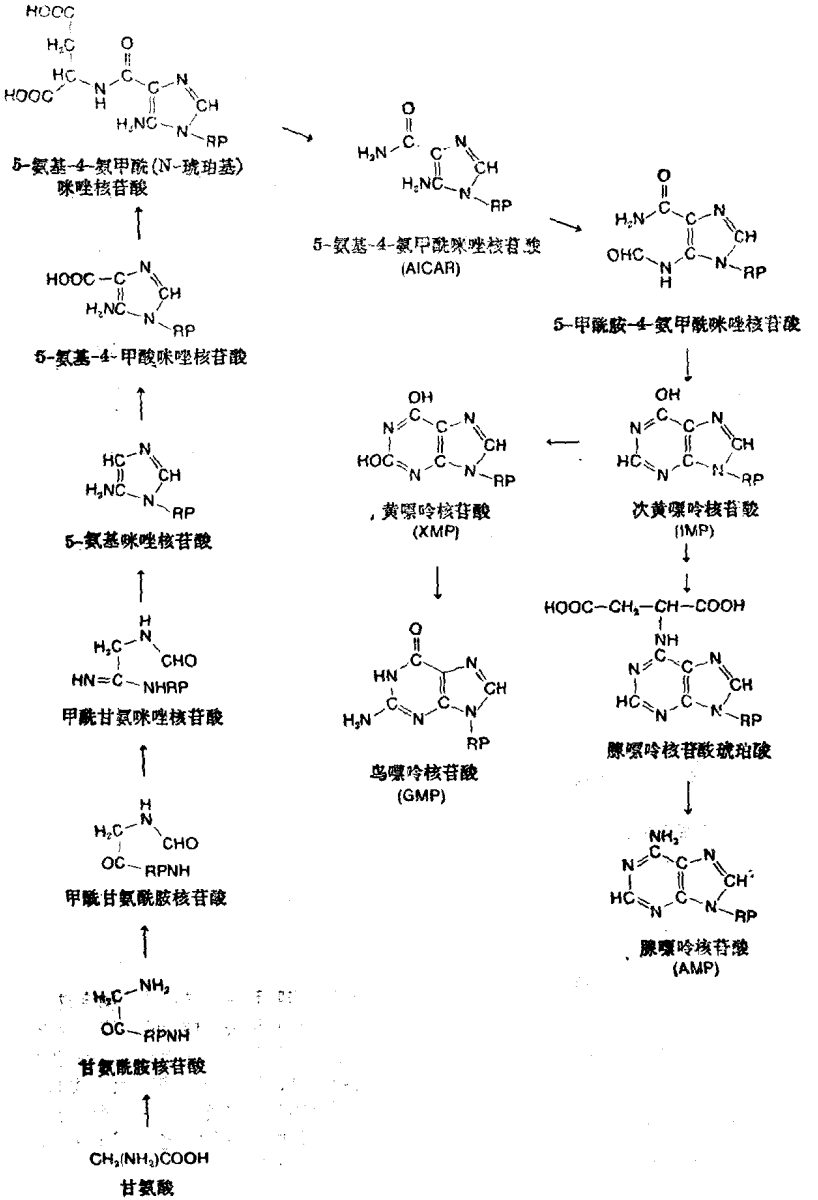

三磷酸核苷酸
脱氧腺苷三磷酸
dATP


脱氧鸟苷三磷酸
dGTP


脱氧胞苷三磷酸
dCTP


脱氧胸苷三磷酸
dTTP

(a)(4)


5-氨基-4-氨甲酰(N-琥珀基)
咪唑核苷酸
5-氨基-4-氨甲酰咪唑核苷酸
(AICAR)
5-甲酰胺-4-氨甲酰咪唑核苷酸
5-氨基-4-甲酸咪唑核苷酸
5-氨基咪唑核苷酸
次黄嘌呤核苷酸
(IMP)
,黄嘌呤核苷酸
(XMP)
甲酰甘氨咪唑核苷酸
腺嘌呤核苷酸琥珀酸
鸟嘌呤核苷酸
(GMP)
甲酰甘氨酰胺核苷酸
腺嘌呤核苷酸
(AMP)
甘氨酰胺核苷酸
甘氨酸

(b)

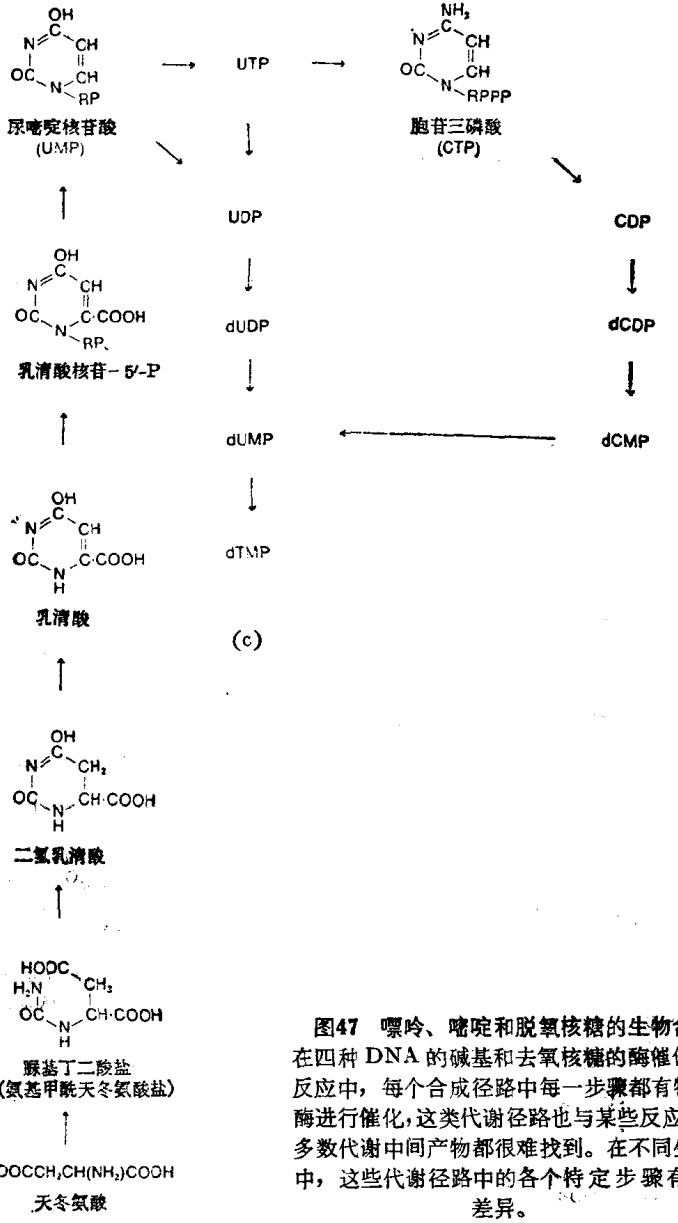


图47　嘌呤、嘧啶和脱氧核糖的生物合成

在四种 DNA 的碱基和去氧核糖的酶催化合成反应中，每个合成径路中每一步骤都有特定的酶进行催化，这类代谢径路也与某些反应类似，多数代谢中间产物都很难找到。在不同生物体中，这些代谢径路中的各个特定步骤有很大差异。

Watson 与 Crick 在他们的 DNA 双螺旋结构模型中，广泛吸收了多种多样的科学资料，其中也包括前面所讨论的遗传实验，Chargaff 的工作以及 Maurice Wilkins 与 Rosalind Franklin 的X射线衍射数据在内。后来 Watson 又写过一本题为 "Double Helix"(«双螺旋»)的小册子，作者对某些历程、科学背景以及为导致双螺旋模型最终得以完成的钻研情况均提出了自己的说明。与之相映成趣的是另一本反映 Franklin 观点的著作[«罗瑟琳·弗兰克林与DNA»(«Rosalind Franklin and DNA»)]，Wilkins 和 Franklin 研究DNA是从其X射线衍射特性的角度出发的。而从X射线特性所得出的令人信服的根据，只不过表明了DNA分子应呈一种螺旋形结构。

大体说来，据X线衍射结果可以推理说明如下：在绝大多数有机化合物分子中键联原子的间距为 1～2Å，非键联原子的间距则为 3～4Å。从理论上，X射线波长甚短，为 1.5Å，足以分辨出处于有机分子内部的各个原子的位置，可是实际情况是，一张X射线衍射图(图 48)所反映出的仅仅只是分子结构中某些周期性的，亦即有规则地出现重复的组分。不管 DNA 材料来自哪一生物种属，其X射线衍射"图象"总会显示出周期性的即结晶形的花样。按照花样推断，DNA 长丝看起来似乎沿其长度的任何一点，在三度空间上(三个方向)其结构组分都是有规则地相间排列的。不过关于认识 DNA 螺旋结构的启示还是来自于这一事实：即在衍射图上的一个结晶图重复花样与 DNA 分子中的一个化学组成重复单位(如从一个磷酸残基到第二个磷酸残基)在长度上出入颇大，例如，从 DNA 分子的化学组成单位来看，从一个磷酸碱基到下一个磷酸碱基的距离为 3.4Å；可是结晶图上的重复部分距离却在 28Å到 34Å之间，视 DNA 分子中的水分子含量而异。这意味着，在一个结晶图的重复部分之内似乎包含有好几个化学组成单位的重复。因之，顺应这一客观资料，最简易明了的解释只能认为 DNA 分子是盘绕式的螺旋结构。这一结构，也是 Watson 和 Crick 所假定的 DNA 分子结构原型之一。图 49 所示是关于 DNA 分子结构的

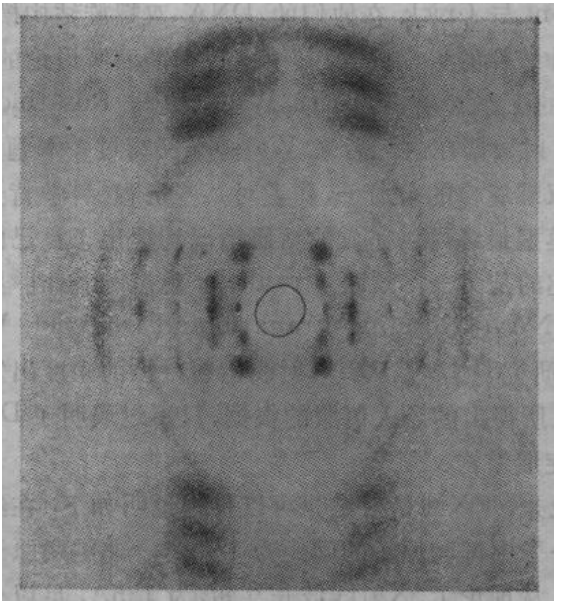


图 48 DNA 分子的 X-衍射图

本图是一张 DNA 分子的早期的X-衍射图，从这类衍射图来看，可以把 DNA 分子设想为是一条细长的纤丝，整个纤丝上的直径幅度大体相同(15～20Å)。(本图由皇家学会会员、伦敦大学的M. H.F. Wilkins 先生赠送。)

一般模型，附图说明中亦列入了几篇详细说明这一模型的重要文献。

在把基因和 DNA 挂起钩来，以及在提供思路，建立 DNA 分子结构的物理构型方面，微生物学、遗传学、生物化学以及物理学等所有学科都作出了自己的贡献。但是，即使各门学科所取得的认识最终归结于双螺旋模型的脱颖而出，人们还是搞不清楚遗传信息从 DNA 向另一类大分子化合物，特别是向蛋白质完成传递的方式。由于 A、T、G、C 是明显地能使一类 DNA 分子有别于另一类 DNA 分子的四种化合物分子，这一事实暗示着或许可能存在有所谓遗传字母表的物质基础。不过知道有遗传字母表还不等于已经掌握了遗传语言，回顾一下这一情景，确实令人感到惊奇：有多少生化遗传学家曾为成功地揭示突变基因与突变酶分子之间所存在的一对一关系而感到激动和鼓舞，却居然对于蛋白质的氨

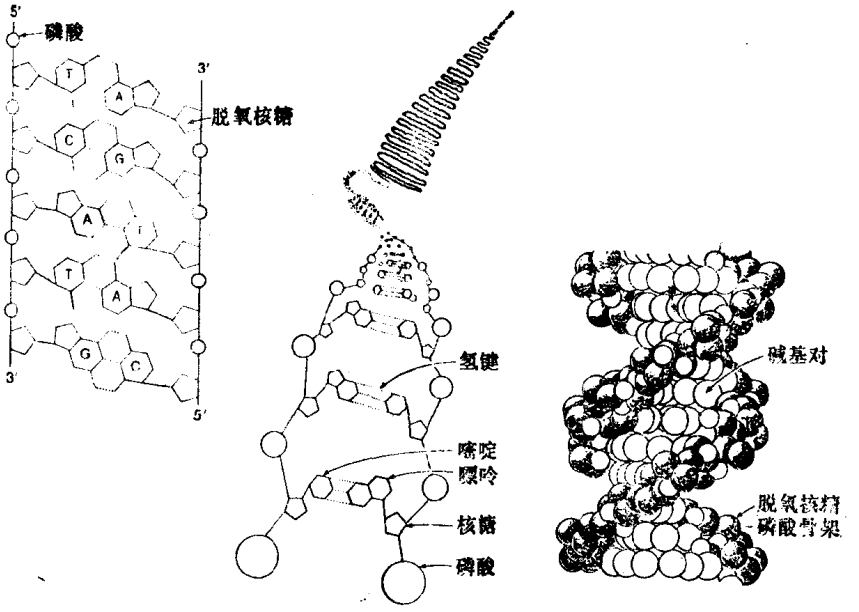
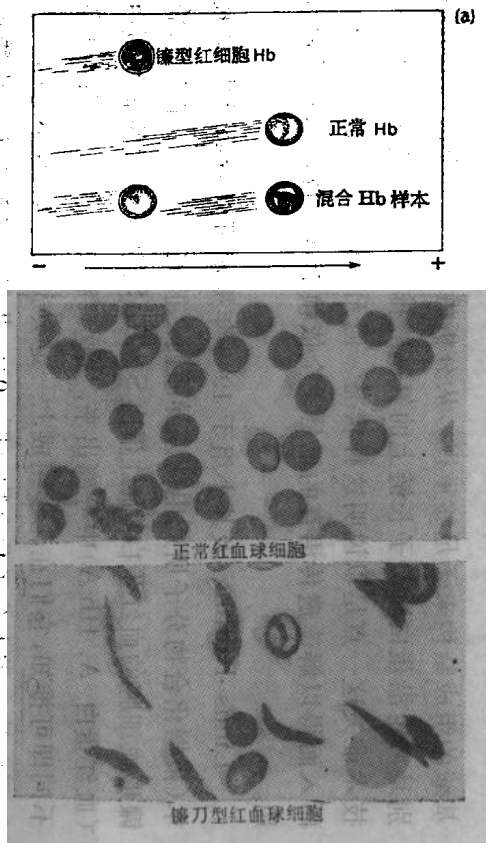


图 49　有关 DNA 结构的较为详细的模型

从本图展示的 DNA 图象，读者可以将 DNA 分子的空间结构与其平面结构作一番对比。关于如何从平面结构转变为"装袋"式结构的过程亦不妨作一番细心的观察（本图仅仅只不过是这一过程的示意模式——顺便提一下，所谓 DNA 装袋还是一个尚未解决的问题）。(参见 Ingram 1965, Kornberg 1974.及 Wilson 1966。)

基酸的组成结构未予足够重视。从分析蛋白质的氨基酸的成分特别是测定沿蛋白质分子长链上的氨基酸排列顺序的问题一直成为技术难关这一点足以说明这种状况。但无论如何，人们终于破门而入而得以解读遗传语言的正是完全依赖于这一技术的发展和运用。

Linus Pauling 及其小组于 1949 年提出可以借助于电泳技术(根据化合物分子所载电荷差异而使之彼此分开的一种技术)将镰形血细胞的血红蛋白(血红蛋白 S, Hb S)从正常型的血红蛋白(血红蛋白 A, Hb A)中分离出来，从这二种血红蛋白所载的电荷不同即可推知，它们在化学组成上一定会表现出差别，不过因为血红蛋白是一类大分子化合物（图 50），共包含有 574 个氨基酸残



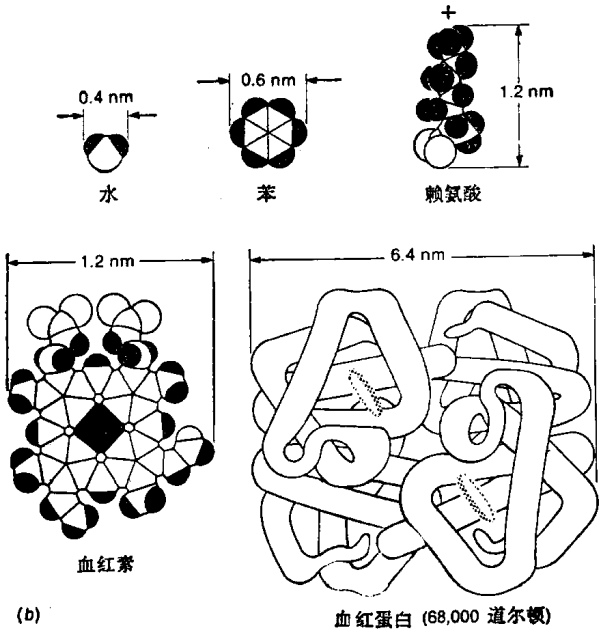


图 50　血红蛋白 A(Hb A)与血红蛋白 S(Hb S)

(a) 在电泳场中 Hb A 与 Hb S 的电泳花样即可简便地测出这两种血红蛋白分子的电泳差异。例如，由于 Hb A 向正极泳动较快，即可测知 Hb A 所带的负电荷高于 Hb S。　(b) 各种化合物分子大小比较示意图，读者由此可获得一种印象：血红蛋白分子可真大啊。

基。这样一种复杂情况给人们在鉴别 Hb S 和 Hb A 化学成分差别的工作中带来了极大困难。然而到了 1956 年，Vernon Ingram 终于提出证据，认为这二种血红蛋白分子，实际上只存在着一个氨基酸残基的差异。

单个蛋白质分子链称为多肽链，多肽链是由多个氨基酸残基彼此以所谓肽键相连聚合而成的长链。血红蛋白分子由四条多肽链构成。四条多肽链分属于两类，其中二条为 α 链，二条为 β 链。Ingram 所发现的氨基酸差异是在 β 链上，也就是说，在 Hb S 与 Hb A 中，各自的 α 链是彼此相同的，在 Hb A 的 β 链上从末端数起，第六个位置的氨基酸为谷氨酸，而在 Hb S 的相应位置上，却是一个缬氨酸(图 51)。在这种血红蛋白的分子中，所发现的二个氨基酸的差别，不仅说明了 Pauling 所揭示的分子电荷的差异，而且也展现了这一事实：即单个基因突变可在蛋白质分子上导致产生一个在该突变基因遗传控制下的单个氨基酸的置换，而这一微不足道的化学成分的改变，却可能会带来巨大的功能与表型上的不同（人们早已知道镰形红血细胞的血红蛋白是一个遗传的单因子性状。当该基因处于纯合态时，由于 Hb S 的功能异常即可使人发病致死)。

Ingram 的工作引起了人们对遗传语言的强烈关注。在这方面似乎可以合理地设想，作为突变最简单形式的基因突变，其结果仅仅只是改变 DNA 分子长链的残基顺序，而 DNA 碱基顺序的变化又随之能转译为与之相应的蛋白质分子中某个氨基酸在排列顺序上的更换（图 52)。到了 1961 年，Charles Yanofsky 及其合作者又令人信服地证明了细菌 DNA 上碱基的变化能够以专一的形式诱导成功，而且根据碱基发生的改变可以预测到可能引起的蛋白质某些氨基酸的改变。

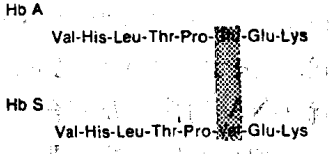


图 51，Hb A 和 Hb S 的氨基酸差别
图中的肽链顺序是 HbA 与 Hb S β 链中的一个片断。从片断顺序可显示出这二种多肽之间的氨基酸差异。Hb A 与 Hb S 的 α 链以及 β 链中，除此片断以外的其余部分，碱基顺序完全相同。

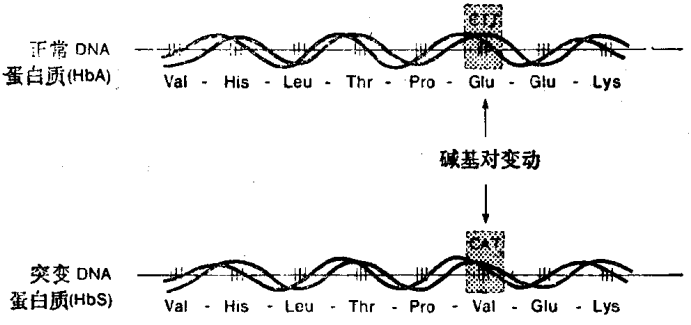


图 52　突变的化学基础

人们从 Ingram 血红蛋白的研究所得出的关于信息流的一般模式是一个基因特异性地规定着一个特定蛋白质中的氨基酸顺序，而一个遗传突变即可幕想为是 DNA 分子中的一个碱基对发生变化，碱基对的变化又进而导致相应的蛋白质中的氨基酸改变。DNA 上的"错误"能够遗传，并由于突变蛋白质功能异常，而使表型发育产生变异。图中所示仅为 DNA 双螺旋中一条核苷酸链的碱基顺序，相应的氨基酸排列顺序即由之进行编码。

本世纪五十年代，又兴起了另外一系列实验的热潮，这便是开展对 DNA 合成机理的研究。Arthur Kornberg 亲自主持了这些实验。作者假设，DNA 合成也是一种酶促反应，并为之曾化了一年时间试图能把想象中的可能存在着的酶分离出来，旨在希望通过对 DNA 合成过程的认识将能够解释作为遗传物质的重要功能之一的复制问题(图 53)。

以后研究 DNA 的历史进程又进入了"破译遗传密码"阶段。在这一阶段 Marshall Nirenberg 于 1960 年发表的工作可作为代表，关于 DNA 研究的历史沿革，我们且说到这里。下面开始拟就 DNA 的性质及其合成机理作一叙述。

如前面所作的简要的历史回顾所表明的，分子遗传学的形成最终还是仰赖于 DNA、RNA 和蛋白质等大分子化合物的结构与功能的研究成果，在本书的其余章节里，我们还要指出，并不是所有的遗传学问题，都已经全部解决了，而且对于分子遗传学来说，它要探讨的问题恐怕比之 DNA⟶RNA⟶蛋白质这一"中心法则"所包含的内容更为广泛。

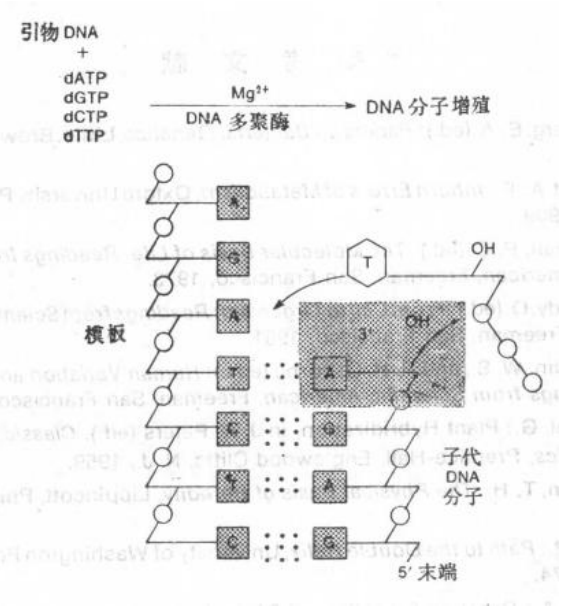


图 53 DNA 复制

Kornberg 最早关于 DNA 的合成工作是以下述前提为根据的：当存在四种三磷酸脱氧核苷酸以及一个"引物 DNA" 或(与)"模板" DNA 分子和 Mg^{++}离子的条件下，一种特定的酶(DNA 聚合酶)便能够催化合成新的DNA分子。

由于分子遗传学本身已经日益与范围更加广泛的分子生物学交织在一起，所以要想把 DNA⟶RNA⟶蛋白质的信息传递有关的种种实验和发现，按照精确的时间顺序，重新作一番归纳和整理是不可能的。评论学家们按各自不同的看法、着重点和解释，把上述种种实验和发展划成主要及次要之分时，其说法并不一样。但是有一点无庸疑义，即过去和现在任何一个从事信息流研究的人，都为上述学科的发展，作出了自己的贡献。门户一旦打开，某个具有远见卓识的学者，可能由之取得下一步的突破只不过是一个时间问题，甚至在取得"信息流"这把宝钥之前，就有不少人从经验直感到要进行科学攻关其入口处就在这里。如果我们从分子遗传学的起源学科(化学、遗传学和进化论)受到启发的话，那末我们就能从大量的实验中选出几个关键性实验：紧紧扣住信息流这一主题不放。

参 考 文 献

Adelberg, E. A. (ed.): *Papers on Bacterial Genetics*, Little, Brown, Boston, 1960.

Garrod, A. E.: *Inborn Errors of Metabolism*, Oxford University Press, London, 1909.

Hanawalt, P. C. (ed.): *The Molecular Basis of Life: Readings from Scientific American*, Freeman, San Francisco, 1973.

Kennedy, D. (ed.): *From Cell to Organism: Readings from Scientific American*, Freeman, San Francisco, 1961.

Laughlin, W. S., and R. H. Osborne (eds.): *Human Variation and Origins: Readings from Scientific American*, Freeman, San Francisco, 1967.

Mendel, G.: Plant Hybridization, in J. A. Peters (ed.), *Classic Papers in Genetics*, Prentice-Hall, Englewood Cliffs, N. J., 1959.

Morgan, T. H.: *The Physical Basis of Heredity*, Lippincott, Philadelphia, 1919.

Olby, R.: *Path to the Double Helix*, University of Washington Press, Seattle; 1974.

Sayre, A.: *Rosalind Franklin and DNA*, Norton, New York, 1975.

Srb, A., R. D. Owen, and R. S. Edgar: *General Genetics*, 2d ed., San Francisco, 1965.

Stent, G. S. (ed.): *Papers on Bacterial Viruses*, Little, Brown, Boston, 1960.

Sturtevant, A. H.: *A History of Genetics*, Harper & Row, New York, 1965.

Watson, J. D.: *The Double Helix*, Atheneum, New York, 1968.

第三章　DNA

复　制

到 1953 年有关遗传物质的问题已经提出如下几点概念：

1．从一个世代到另一世代，染色体具有高度稳定性。

2．自然突变有其本身内在的规律性，突变频率一般较低，人工的方法能够提高机体突变频率。

3．突变基因和突变型的代谢途径（酶）之间存在着一对一的对应关系。

4．在同源染色体之间，重组可以发生在基因之间或是基因内部。

5．遗传物质是 DNA。

6．DNA 大分子呈双螺旋结构。

遗传信息的语言在没有被人们破译之前，对于沿着染色体长度的基因如何分段，以及其他大分子化合物在基因的结构和功能方面是否也与 DNA 协调作用的问题可以说一无线索可寻。

在 1953 年以前，对基因从生物学的角度开展了大量工作，但是在化学性质方面的研究则相对较少。在 DNA 结构模型提出以后，人们对基因结构和功能的化学本质问题兴趣倍增；分子遗传学亦从而进入了它发展的"黄金时代"。双螺旋结构含义十分深广，越来越多的"开拓者"将可以在这一概念的基础上不断提出问题，施展自己的才华，但即使如此，首先突出的还是如何把这一结构模型与经典的遗传学理论衔接起来。用历史的尺度去衡量，可以看出这里所讨论的问题，有不少仅是早期遗传学实验中所出现问题的

某种延伸而已。例如，有些人一直在集中研究 DNA 结构与生长现象、细胞分裂和遗传物质增幅之间的关系。更准确地说，他们提出的课题是：在"细胞"分裂以前，诸如病毒、细菌、果蝇等等染色体究竟是怎样进行复制的？

有相当多的科学家指出，必须要去寻找 DNA 结构和突变机制之间的关系。Watson 和 Crick 在他们的早期论文中曾就DNA 分子里一对碱基对由 DNA 分子中另一些碱基对所替换的理论机制进行了讨论，作者认为，这样一种置换的机理可用来解释某些遗传突变的起因。还有人在 DNA 结构基础上着手于遗传重组的结构模型和交换的结构模型的研究，从而使遗传重组的问题又周期性地成为人们研究的主题，这在遗传学的整个历史中几乎是屡见不鲜的。遗传物质的新的分子模型在为重新研究过去的遗传模型和理论以及建立全新的模型和理论方面，也是一股新的动力。

本章讨论的内容是涉及有关 DNA 复制、突变和遗传重组的某些分子遗传学的思想方法和实验。从一些书籍以及这一题目的种种讨论中，可以得知由于 1953 年 DNA 结构模型的建立而掀起的一股研究高潮大都是起因于这样一种期望：对 DNA 结构模型的真实性进行验测。不过从另一方面来看，可能这样一种说法更加接近于事实：即所提出的 DNA 结构模型对于促使一些人放下他们现有的工作，也许是一种言之成理的依据，因为应用这一模型理论作为指导思想将会开拓出一个更富于成果的全新领域。总而言之，DNA 双螺旋结构模型的问世，使过去有关在复制、突变和重组等诸多问题上都面临着新的挑战。

人们按照 DNA 结构模型首先据以推理和加以说明的便是遗传物质的复制机理，这是十分合理的。从单个染色体进行分裂和增殖，产生两个彼此完全相同并且与亲本染色体如出一辙的子代染色体，与此同时，基因经过如此传递能够数百代难免差错而不受损害。长期以来人们对于保证复制精确性的这一机制一直存在争议。现在有了 DNA 结构模型，毫无疑问也就找到了认识这一机理本质的最初的一些线索。在生物学领域中象 DNA 研究那样简

便而精密地表明出结构和功能之间关系者确实是为数无几的。

从生物学角度来看，DNA 结构中突出的特征是它所固有的双链性质。双链之间由氢键相连，碱基 A-T 对之间具有二条氢键，碱基 G-C 对之间具有三条氢键。按照这个模型推论，假如所有氢键在所有碱基对之间全部断裂，原先 DNA 的两条链(半条分子)将会相互分离。氢键比之共价键，吸引力相对较弱，因而可以预

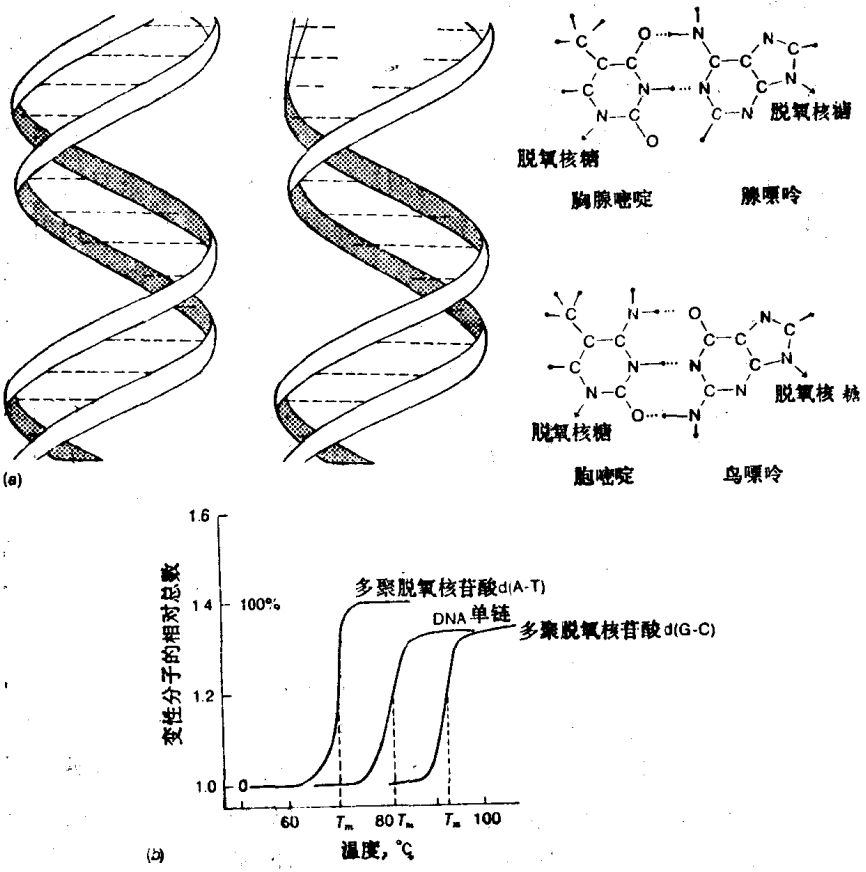


图 54　氢键和 DNA 结构

(a) 在 DNA 分子的互补链之间，碱基对A-T和G-C之间由氢键相连结。

(b) 用热处理使 DNA 降解时，双链 DNA 分子即降解为单链 DNA 分子。从变性-温度曲线可知 G-C 碱基对一般比 A-T 碱基对降解时需要更多的热量。

计，在应力作用下首先断裂的将是氢键，而不是共价键(图 54)。亦即当通过增加能量使 DNA 双链降解时，一当 DNA 双链中的氢键断裂以后，DNA 双链分子随即降解为二条单链。

所有这一切都不难在实验室里获得证明，将溶液中的 DNA 分子加热至近 70℃时，双链分子的 DNA 即变性离解成单链。在低能量作用下，迄今没有任何共价键被切断的证据，这意味着 DNA 双链分子并非是通过共价键相连结在一起的。

需要说明的是双螺旋链之间的氢键具有一种强制性性能，它使较大的嘌呤碱同较小的嘧啶碱配对在一起（同时亦使特定的嘌呤碱必须和特定的嘧啶碱配对——译注）(图 54)，以保证两条核糖磷酸骨架在空间上有均匀的构型。这样就消除了整条双链在不同链段上可能产生的形状差异以及为维持两条单链相连而提出的不同能量需求。既然维持二条长链结合在一起的氢键数目很大，并且沿着整条的 DNA 长链，各个氢键的空间结构极为规则；所以在不存在氢键大量断裂的情况下，要使一个氢键或少数几个氢键发生断裂并不容易。这一点十分重要，其次因为从整个分子的外形来看，嘌呤和嘧啶对的位置排列相当紧密，即使有几个氢键偶尔断裂，也会重新产生新键以完成修补。基于这一理由可见，那怕保持双链相连的个别键力量不强，但是 DNA 分子双螺旋结构仍是一个相当稳定的构型。

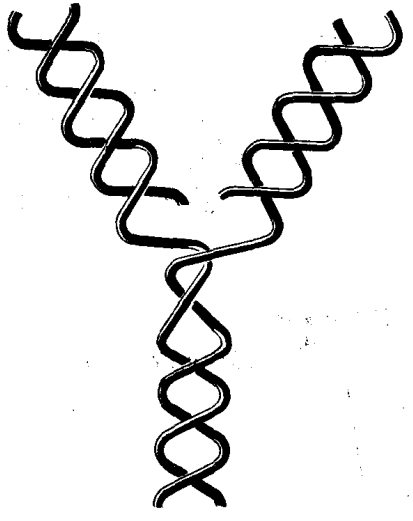


图 55 DNA 分子复制模型

模型表示复制前原先的 DNA 分子中的二条链相互拆开，再以每一条拆开的单链作为模板，合成新的子链。由此可以预期合成完毕后，新的 DNA 分子必然与原先的 DNA 分子微妙微肖，如出一辙。但本模型不能解释双螺旋的解旋机理和DNA 的合成机理。

考虑到 DNA 分子双螺旋结构的这些基本化学性质，人们不禁要问，这些化学特性在生物学上具有什么意义，二者的关系究竟如何，这一点十分重要。因为最初关于 DNA 分子复制机制的设想就已暗示，这一复制的过程必须以 DNA 分子双链的分开作为起点，DNA 分子双链一经分离，二条链中的每一条链便都成为二条新的子链合成的模板(图 55)。根据这一设想推断，在 DNA 分子合成期间，最初的双螺旋不再保持完整，但是双链中的每一条链却自始至终巍然不动。换句话说，在复制期间原初的 DNA 分子不是全部保留，而是只有其中的一半分子在复制中被原封不动地保存下来。亦即属于所谓的"半保留"复制(图 56)。

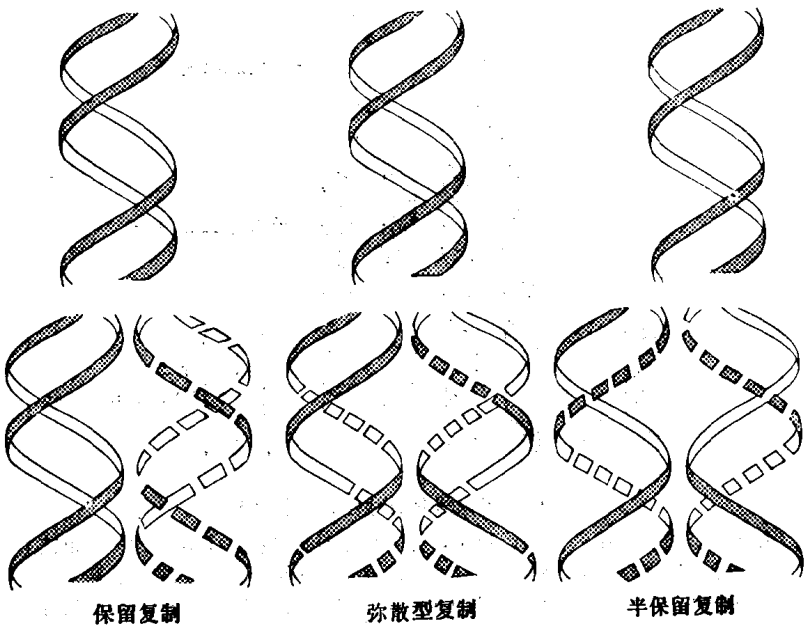


图 56 关于保留复制，半保留复制和弥散型复制的模型

这些模型表明在复制过程中，亲本 DNA 至少有三种不同的去向：

(1) 以完整的双螺旋分子形态保存下来(保留复制)；

(2) 在不同的子代 DNA 分子中，原封不动地保存着亲代 DNA 双链中的一条单链(半保留复制)；

(3) 在子代 DNA 分子中间，亲代 DNA 与子代 DNA 散杂混合在一起(弥散复制)。

图中虚链表示新合成的子代 DNA 链，实链表示亲本链。

假如 DNA 复制是半保留复制(图 57)，可以预见，在经历一个复制周期以后，子代 DNA 分子的两条链中将有一条为完整的原初的亲代 DNA 分子链中的单链，再继续复制到第二代或第三代，随着世代的增加，新的 DNA 分子也越来越多，其中大多数新的 DNA 分子都各具有全新的二条新链，但是两条亲本链并不湮没，它们将分别存留于分子群体中某二条子代 DNA 分子之内。

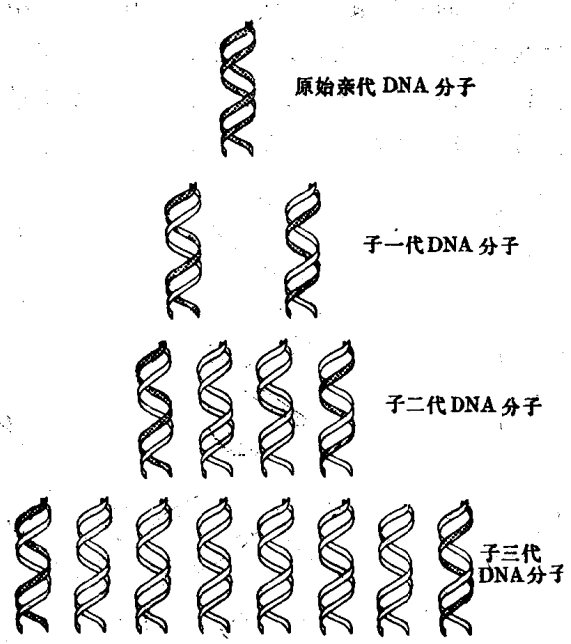


图 57　半保留复制过程的二个世代

复制一个周期，每一个子代 DNA 分子将分别具有亲代 DNA 的一条单链和一条新合成的子链。复制第二代以后，子代 DNA 分子有二种类型：四个子代 DNA 分子中有二个子代 DNA 分子与子一代 DNA 分子相同，即具有一条亲本单链和一条子代新链，另外二个 DNA 分子则全由二条新链组成。

在这一设想提出之后不久，人们即着手实验验证，这里我们介绍二个根据上述设想而设计的实验，其中仅有一个实验"解决"了问题，传诵至今，另一个实验却是作为不太成功的记录载入历史的。

第一个实验是 Cyrus Levinthal 和 Charles Thomas 的工

作，他们将^{32}P〔一种因 β 粒子逸失而衰变的磷的同位素（表 3）〕渗入到 T_2 噬菌体的 DNA 分子里面，旨在追踪该同位素在经历几轮 DNA 复制后的去向。在无放射性元素的培养液里，原先整个被标记的 DNA 双螺旋分子在复制经过第一轮之后，预期会产生二个各有一半被标记的子代 DNA 分子。也就是说，原初被标记的链将保留不变，而新合成的子代链没有标记。经过第二轮复制，所得的子代 DNA 分子将有二种类型；其中半数与子一代 DNA 分子相同，即有一半分子被标记，另一半分子不被标记。另一类是子二代 DNA 分子完全没有放射性标志。从这一模型看来，这一实验本身似乎是无懈可击的（图 58）。

表 3　放射性同位素释放粒子的类型

粒　子　类　型	特　性
α	氦核，俱有二个中子和二个质子，直线前进（由于质量大）
	穿透力较低（对健康危害性亦低）
	能完全被金属或玻璃容器壁吸收
β	正电子或负电子
	负电子质量＝9×10^{-28}g
	能够穿透细胞核（对健康危害性严重）
	能够穿透铝板
	大部分能被厚玻璃或金属容器壁吸收
γ	为一类电磁波，波长在紫外光以下
	主要为光子，有较大的质量，不带电荷，能穿透厚金属板
	贮藏时必须置于用厚铅板为壁的特置容器内

T_2 噬菌体是侵染大肠杆菌细胞的噬菌体（图 59）。其染色体为单个的 DNA 分子。一当噬菌体染色体注入到细菌的细胞里，即复制，产生二个 DNA 分子。二个 DNA 分子再复制成四个 DNA 分子，照此办理，直到形成 200 多个噬菌体染色体。噬菌体颗粒成熟后，便开始溶菌，释放出新的噬菌体颗粒。假如从侵入到

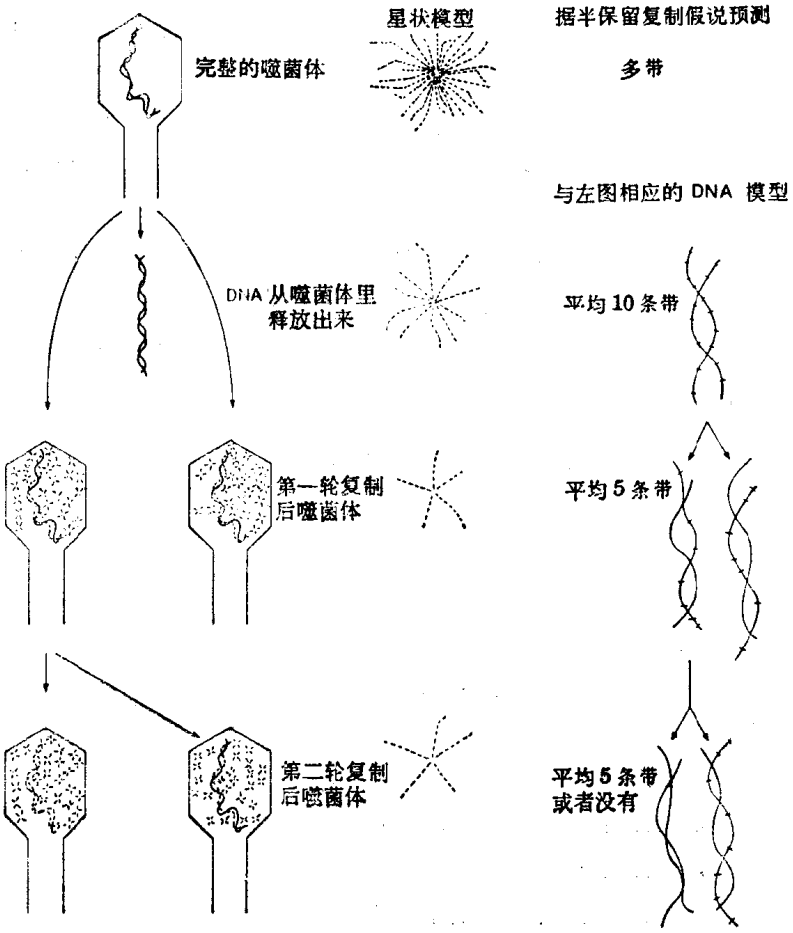


图 58　标记实验设计程序

点号表示 ^{32}P 原子的蜕变，在标记完全的噬菌体里 ^{32}P 的密度是很高的，如图所示，预期随着半保留复制有规律地进行，^{32}P 的密度亦就有规律地消减。

释放出新的 T_2 噬菌体这一轮生活周期，开始时只有一条噬菌体染色体参与其事，在形成的诸多子代噬菌体 T_2 中，总共只会有两个噬菌体的染色体各自包括有一条原初亲本的链。也就是说对于有标志的亲本原初材料来讲已接近稀释 200 倍了（图 60）。在细菌细胞是被多个噬菌体感染的情况下（每个感染的噬菌体 DNA 分

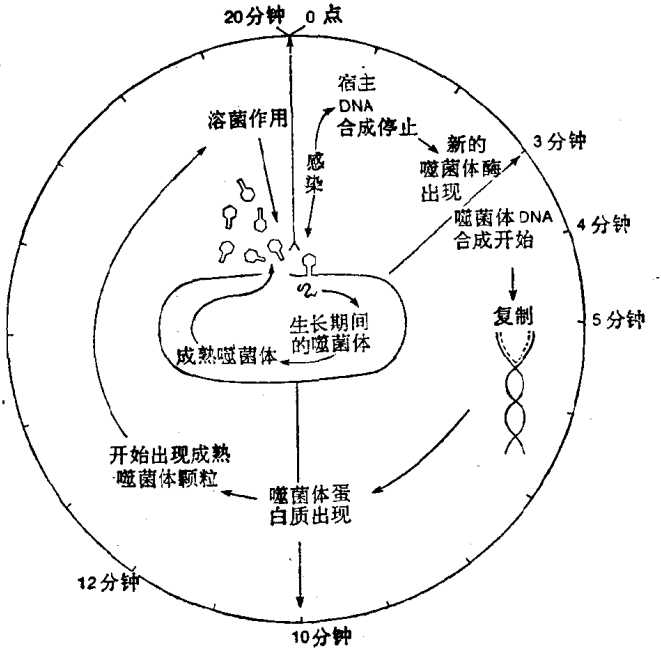


图 59 T_2 噬菌体的生活周期

圆圈表示 T_2 噬菌体一个世代周期的时间为 20 分钟。从噬菌体颗粒的 DNA 感染细菌细胞到溶菌大约需要 20 分钟，约产生 200 个新颗粒。T_2 噬菌体 DNA 合成一般于感染 3 分钟后开始，感染 12 分钟后即能见到成熟的 T_2 粒子出现。以上所说，是指 T_2 噬菌体在条件最适宜时的生长情况。

子都作有标志，稀释效应将会相应降低，此即称为多重感染(high multiplicity of infection)。此时感染的亲代噬菌体染色体 DNA 分子愈多，亲代 DNA 分子在溶菌前的稀释程度亦愈低。

有时候在溶菌以前，寄主细胞即罹病死灭，在此种场合，由于死寄主细胞中已经复制一、二次的噬菌体染色体并未破门而出，所以新产生的染色体也就无法实验收集；换言之，即在噬菌体同步生长中，不论新的噬菌体染色体何时成熟，也不可能找到那些被囚禁在死寄主细胞中的新合成的 DNA 分子。可是另一方面，侵染寄主细胞的亲本链却可借助于其放射性标志，予以检测。这样一来，总的稀释效应便大大降低了。

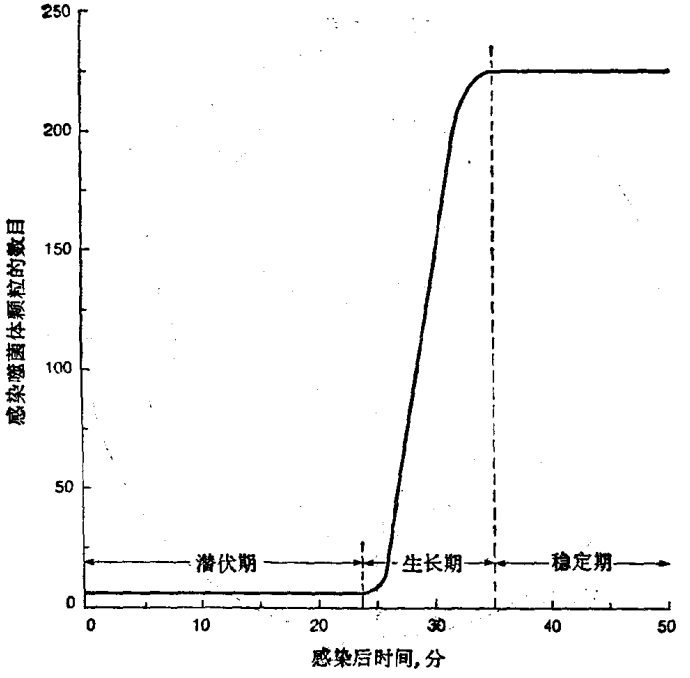


图 60 一个 T_2 噬菌体的阶段生长曲线

现在假定只有一个噬菌体粒子的 DNA 分子上有标记，并感染一个细菌细胞，将可预期在经历一个噬菌体繁殖周期之后原初的 DNA 分子将被稀释 200 倍以上。换句话说，也就是在 200 个噬菌体颗粒中仅有二个噬菌体颗粒含有一条初始 DNA 分子的单链(有标志，其他的噬菌体颗粒所含的 DNA 分子全部都是无标志双链)。

将从溶菌前的大肠杆菌细胞或溶菌后释放的噬菌体颗粒所提取的 DNA 分子进行喷镀照相，便可以像一般照片一样，录下来自 β 射线的放射性位点。从而，即可根据衰变中的磷原子进一步准确定位(图 61)。从图中可见，有为数可观的 DNA 分子都含有放射性磷原子，它们乃是原初亲本噬菌体中所"标志"的 DNA 分子链。从放射位点的数目来推算每个 DNA 分子中 ^{32}P 的原子数目确实是一种行之有效的方法。同样亦可将那些具有单链放射性的 DNA 分子的数目与双链完全标记的 DNA 分子数目进行对比，假如二者之间的数目比例，亦即具有半放射性的子代 DNA 分子与

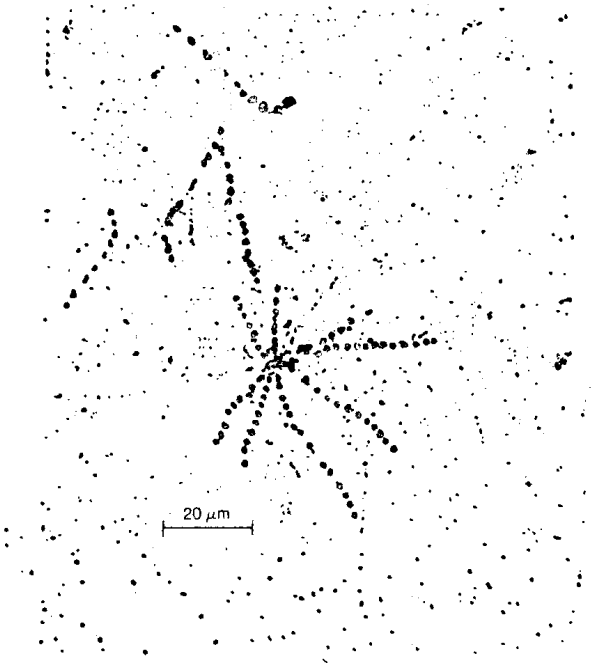


图 61　用 ^{32}P 标记的 DNA 分子经过几个复制周期后所摄的电镜图片

因为 DNA 分子系处于三维空间的培养基上，从而摄影时只能拍摄“一个角度”，因而在图片中只能看到为数不多的放射性 ^{32}P 原子，通过上下调节焦矩，一般能够拍到个别大致完整的具有标志的DNA分子图片。(由 C. Levinthal 提供。)

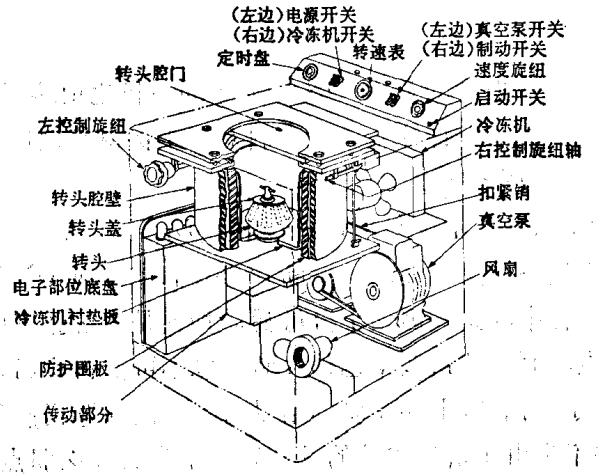
具有全分子放射性的亲代 DNA 分子的数目比值接近于 2 时（指经过一轮复制——译注），这一资料对于半保留复制模型将是一个有力的佐证(图 57)。

图 61 暗示有不少DNA分子相当短小，长度不一，看来它们大概是整个噬菌体染色体的一些断片。人们还发现 DNA 长分子对于提取过程中的某些因素十分敏感，即使是中等强度的切力也会使之断裂，而且从这些染色体片断一般不可能再重新构成原先的 DNA 分子。因此，上述实验的结果对于解决问题仍然是含糊不清的，这并不是由于指导实验的设计思想本身有什么弱点，而是因为分离提取 DNA 的技术存在问题，在提取过程中，原初的 DNA 分

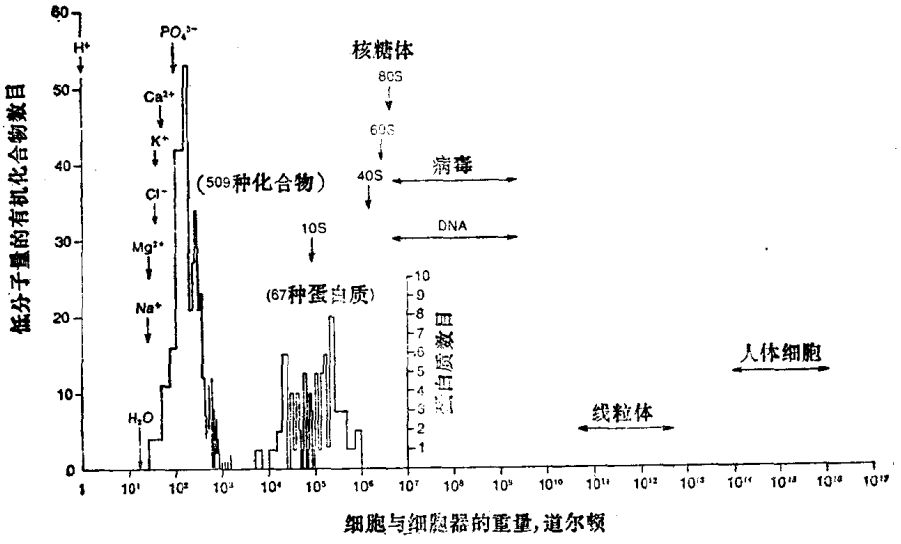
子往往断成数片，从而影响数据统计的可靠性。由于所采用的实验技术不当，结果使一个良好的试验设计受到影响。这在科学史上也许并不是独一无二的孤例。

第二个实验，结果比较明确，虽然其基本概念本质上与第一个实验不相上下。Matthew Meselson 和 Franklin Stahl 所做的这一实验，材料选用细菌的 DNA，而不是噬菌体 DNA。这是因为只要同步复制一、二个周期或是有时在第三周期之后，即可采集细菌细胞供试验取用。另外一个特点是，实验用的标志细菌 DNA 不用 ^{32}P，而使用的是 ^{15}N。因为 ^{15}N 不是放射性同位素，故不能通过放射性蜕变进行检测，不过 ^{15}N 是比 ^{14}N 较重的同位素，用 ^{15}N 标记的 DNA 分子同样亦比用 ^{14}N 标记的 DNA 分子要重（密度大些）。

重的 DNA 分子借助于超速离心机可将其从轻的 DNA 分子中分离出来，其基本原理可参见图 62 所示。Meselson-Stahl 实验的要点在于将诸如像 DNA 这样的大分子化合物能够在与其密度大体相近的介质溶液中长时间超速离心，在强大的离心力的作用下，使溶液从上至下形成某种密度梯度，密度梯度的最低点在离心



(a) L 型超速离心机模型的透视示意图



(b)

图 62　超速离心机工作原理

(a) 从 L 型超速离心机的设计中可以看到，一个固定的转体芯子，芯子直轴与离心管成一定角度。假如放置 DNA 样品的溶液密度与样品 DNA 相近，经超速离心数小时后，在重力作用下便会产生某种密度梯度。此时，DNA 分子在沉降位置上即达到一动态平衡，亦即某一级 DNA 的密度与之所处位置的介质溶液的密度相同。从而导致不同密度的DNA分子分别处于密度梯度等级的不同区段。(引自 D. J. Weber, 1973。)

(b) 本图表示了各种大分子、细胞器、以及细胞之间的相对体形大小，以及用沉降常数（沉降速率的数值）所标志的分子量比较（道尔顿）。(引自 N. G. Anderson, CRC 生物化学手册，1968。)

管顶端，最高密度则位于离心管底部，各个密度区段都与各区段的离心力强度相当，亦即上部较弱，下部较强。在溶液里每一级梯度上的各种大小的 DNA 分子最终会处于各不相同的梯度节段，而每一节段的介质溶液的密度此时亦与处于其中的 DNA 分子的密度相等。不同密度的较轻型分子因此通过这一过程，仅管各部分 DNA 分子的密度差异极其微小，也能够人工将之各各拆离。就各类 DNA 分子组分所处的区段到梯度等级上端的距离，测量之后，便可估算出它们各自的密度大小。

细菌在含有 ^{15}N 标记元素的培养液中培养二个或二个以上世

代以后，其染色体 DNA 即可全部标记成功，然后将被标志的细菌再转到含有 ^{14}N 的培养液里，培养一个、二个或更多的世代。按所设计的不同培养时间取样，提取 DNA。在氯化铯(CsCl)溶液里，进行密度梯度离心，图 63 所示即为达到梯度平衡时的实验的结果。从图 63 可见，全标志细胞的 DNA 为“重 DNA 分子”，复制

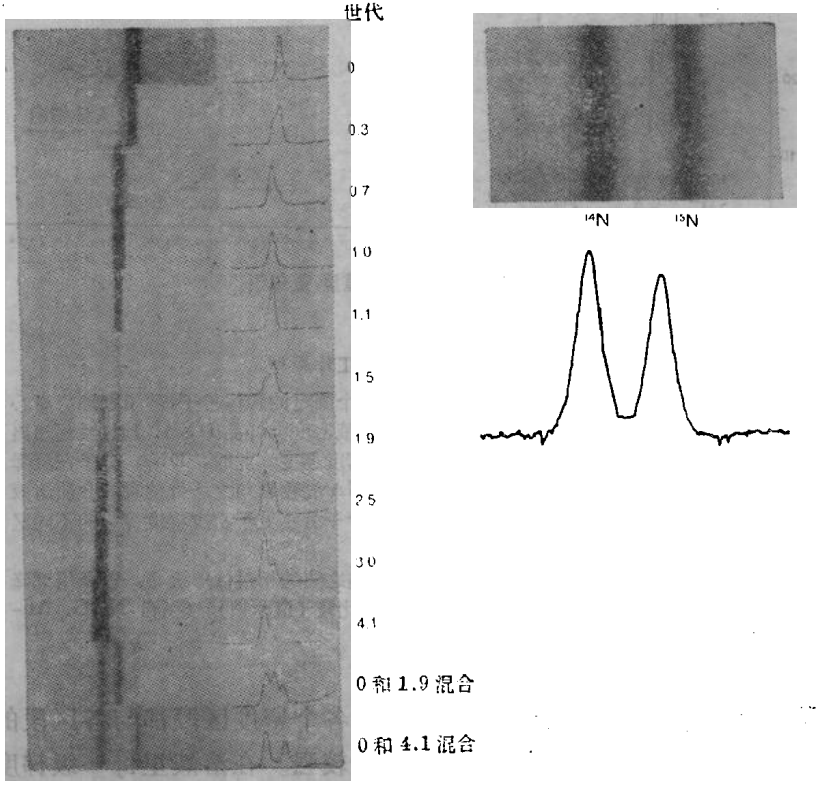


图 63 Meselson-Stahl 的实验结果

左侧谱带是 DNA带的紫外光吸收光谱的照片。氯化铯($CsCl_2$)密度梯度离心 20 小时，转速 45,000 rpm。右面的曲线峰是相应 DNA 带，用密度计测得的 DNA含量变化。取样的世代时间以全标志细胞置入^{14}N 培养液中生长的时间计算。计算全标志细胞刚刚置入时(即零时)，细胞全部为重 DNA分子；复制一代以后，即变为中等密度的DNA分子，复制二个世代以后，此时的 DNA有二种：一半为中密度 DNA，一半为轻密度 DNA。在其后的世代中，中密度 DNA 进一步稀释，但没有任何其他密度级的 DNA 出现。全部实验一共只观察到重、中、轻这三种密度级的 DNA分子种类。(引自 M. Meselson 与 F. W. Stahl, PNAS USA, 43: 581, 1958。)

一代的细胞(即第一代)变为中等密度的 DNA 分子，第二代细胞则具有二种类型的 DNA 分子:一半为中等密度 DNA，一半为轻 DNA。这些资料对于半保留复制模型来说，无疑都是令人信服的佐证。

从关于 DNA 合成半保留复制模型的这两个实验，可以看到实验技术的分辨力对于实验本身的成败是何等的重要。在实验科学中，建立一个杰出的模型往往是很有价值的，但是任何一个成功的模型却又决非尽善尽美。这也何尝不与实验手段有关。此外，这些工作还说明了另一个事实，即实验技术的发展与进步会进一步带动理论、模型和观念方面的新的突破。这一事实对于那些把主要注意力集中于直接研究科学规律本身，而忽视科学研究技术与手段的人来说，不仅是一种挑战，而且常常使后者在工作中陷入进退维谷的困难处境。

本世纪五十年代,DNA的半保留复制模型得到普遍的确认，但是知道了 DNA 半保留复制的机理并不等于已经弄清楚 DNA 复制的具体进程。人们对于 DNA 究竟是如何进行复制的问题，依然一无所知。半保留复制暗示，双螺旋 DNA 分子中的每一条单链是作为新链合成的指导者起作用的。但是，我们不明白，新链的合成是不是酶促反应，例如，还有亲本链是作为子代链拷贝的模板呢，还是仅仅只是作为反应进程的某种引子，等等。为了揭示复制机制的有关生化过程的细节，这里需要做大量的各种各样的实验。在发现这类全新的研究方面，首屈一指的当推 Kornberg 小组的工作。他们首先解决的难题是设计一种能据以直接分析 DNA 合成的实验条件。这一工作是以假定 DNA 合成为酶促反应，而且反应有关的酶能够人工分离为其前提的。但在实际上，这两个假说在工作开始时纯系揣想，根本不存在任何实验证据。

从多方面看，研究酶的生物化学乃是一项极为精巧而高超的技艺。这项技艺包括酶分离技术:在对酶进行测试之前，首先得把酶分子从其他细胞的内含物分离提纯。一旦达到相当的纯度，便可以使用简单的生物化学方法来测定其催化特性。不过应该指出，

在对一个可疑的，迄今仍属未知数的酶进行研究分析时，这方面并没有什么现成的方法或技术一定能保证马到成功。在自然科学中有关酶化学的命题研究起步较晚。为了使读者了解 DNA 多聚酶发现过程中所出现的一些问题，本节拟就此略述数端(一般能够催化合成长链多聚物，或能够催化合成诸如 DNA、RNA 这样的多聚类似物的酶称为多聚酶)。

按早期的说法，酶是一种蛋白质，种类极其繁多，每一种酶各以其特定的催化反应为其特征。有些细胞，含酶不多，不过十来种而已，有些细胞，则可能包含有成千上万种酶。最复杂者，如哺乳动物的细胞，无疑具有能合成上千种酶的遗传潜力，虽然由于细胞分化，在一个已特异化的细胞里，并不是全部的基因都能够表达。然而，即使在某些含酶种类为数不多的细胞，也往往同时含有许多其他种蛋白质，因此，从整个的细胞中，要求就某一特定的酶(也许只有几个分子的含量)进行提纯分离，其困难程度，恐怕怎么说也不过分。一般的步骤是：细胞研碎后，用各种方法使之匀浆，将细胞内含物筛分为颗粒和分子二级。在筛分提纯的每一环节，都必须要将所取得的每一组分就所分析的某一特定的酶作比活力测定。

只要一个酶的特性被成功地鉴别出来，测定酶的活性或具体作用便相对比较容易，但是，如果考虑到这一事实，活细胞中的酶(体内系统)经过超长时间的提纯精制，酶活力必定受到损害，而且酶分子本身又只能在一定范围的温度、pH、底物浓度中起作用。因而不难想像，要为试管中的酶制剂人为创造一个优良的离体环境系统，将是何等困难。首先，在用细胞提取物作为酶的来源时，细胞中的所有未知成分便都成了反应物的一部份，这里包括有底物、有毒物质、可能还有一些能够使最终产物被测试以前就会遭到降解的酶等等。总之，要建立一个适当的离体酶检测系统往往都要经过实验—失败—再实验这样长期的反复过程，才有可能取得成功。即使对于技术熟练的人来说也毫无例外(表 4)。

DNA 多聚酶的故事是一个成功的记录，但是却也是一个充满

表 4　Kornberg 纯化 DNA 多聚酶(Pol I)的步骤*

纯　化　步　骤**	酶比活力,单位/mg 总蛋白	产量%(回收率)
1. 提取	3.5	(100)
2. 链霉素	23	47
3. 自溶	170	40
4. 硫酸铵	250	36
	650	
5. EDAE-纤维素	750	36
6. 磷酸纤维素	10,000	28
7. Sephadex G-100	18,000	22

* 引自 Kornberg,1974 年。

** 纯化步骤说明:

1 = 细胞研碎。

2 = 用链霉素处理,因各类大分子化合物在该溶液中溶解度不同,故可据此盐析分离。

3 = 用核酸酶和蛋白酶处理,去除 DNA 和蛋白质。

4 = 在硫酸铵溶液中盐析,原理同 2。

5 = 由于蛋白质的带电基团与 DEAE-纤维素的亲合力不同,故可使用后者分离有关成分。

6 = DNA 多聚酶对于磷酸纤维素有特殊的亲合性(据推测,因为后者和 DNA 分子颇为相似)。

7 = 根据分子大小,提取 DNA 多聚酶(筛滤效应)。

着挫折和失败,不断试验—失败—再试验的曲折反复的过程。例如,早期出现的问题之一是因为细胞既携带着 DNA 多聚酶,又同时含有所谓的核酸酶。而核酸酶降解 DNA 分子的速度几乎与前者合成 DNA 的速度相同,这一矛盾现象应该作何解释呢?

现将最初测试 DNA 多聚酶的情况概括如图 53 所示。底物为脱氧核苷三磷酸(dATP, dGTP, dCTP 和 dTTP),还有二价阳离子(Mg^{2+})和少数 DNA 的分子(按照 DNA 复制模型的推想,它们是作为引物、模板起作用的)。在作为反应混合物的组分使用之前,这四种脱氧核苷三磷酸,先用^{32}P在其三个磷酸根的第一位磷上予以放射性标记。这样,由之形成的(合成的)后代 DNA 分子即可

因其含有放射性磷而能与原初的作为引物或模板加入的亲本 DNA 分子相区别。反应终止后（此时合成 DNA 分子的前体物已被完全耗尽），即从试验溶液中把 DNA 分离出来，并检测其放射性(用三氯乙酸处理，使 DNA 沉淀，再用离心法分离提取)。

作为一种常规方法，在设计检测某个酶的具体作用时，一般都是把试验混合物(这里包括有酶液和底物)分为几个处理，每一个试验处理人为地少放一种底物，然后，与具有全部底物的完全试验物的催化反应物进行对比。这样做对于认识酶的具体作用，测量每一种可能的底物在反应中的重要性是十分合宜的。从这样的试验，结果十分明显，凡是不置入 DNA 作为引物的处理，DNA 合成过程便一律不能进行。从生物学观点来看 DNA 引物的作用是颇为耐人寻味的。因此，早期研究工作的着眼点之一便是要弄清楚在 DNA 分子合成中，亲本 DNA 分子究竟是充当引物的角式呢，还是起某种模板的作用。于是人们又从头做了大量富有生气的实验：通过比较分析亲本 DNA 分子和后代 DNA 分子的碱基排列顺序亦即 DNA 分子的初级结构来确定初始作为底物的 DNA的确切作用。假如亲本 DNA 分子只是合成反应的引物，那沿着后代 DNA 分子长轴方向碱基排列顺序便未必与亲代 DNA 分子的碱基顺序一致。反之，如果亲本 DNA 分子确为新链合成的模板，则可以预期后代 DNA 分子必定会具有亲本 DNA 分子如出一辙的碱基顺序。实际实验结果表明，底物 DNA 的作用完全属于后一种情况。

对于这一设想人们曾产生过疑问，这主要是由于可资说明的证据仅仅只限于核酸离体合成的条件，因之不能排除后代 DNA 与亲代 DNA 的相似性(专一性合成)可能正是离体系统这一特定条件下的产物。后来 Severo Ochoa 在研究 RNA 的合成时，用二磷酸核苷作为多核苷磷酸化酶的底物，结果在同样的离体系统中观察到 RNA 的合成都是非专一性的。也就是说，在这样的合成系统中，它毋需预先加入 RNA 分子作为引物或模板，所合成的 RNA 分子，其碱基顺序同样也是五花八门、杂乱无章。既然 Ochoa

的 RNA 合成系统没有使用 RNA 作为引物，所以 RNA 的非专一性合成应该说是顺理成章的事情。由此可见，DNA 合成具有与此完全不同的另一种体制，在那里作为模板的 DNA 分子是不可或缺的。

所有的 DNA 分子在总体结构上大体不相上下，几乎所有的单链骨架都是由一个脱氧核糖分子和一个磷酸分子交替相间排列组合而成，它们具有 4 种完全相同的含氮碱基 A，T，G 和 C。但是各种 DNA 分子中的 AT 与 GC 比率则有所不同。即使二个分子的 AT 与 GC 比率相同，也没有充分理由认为其初级结构彼此相似。把二个分子的碱基顺序作为判定它们差别的依据要比 AT 与 GC 比率来得可靠。但是现在要问，究竟该如何才能有效地确定仅由四个核苷酸形成的 DNA 分子链中成千上万个核苷酸的碱基顺序呢？课题的难度极高，甚至当需要确定的 DNA 分子为数不大时也是如此。最早用来测定 DNA 分子碱基顺序的定位方法称为最近邻分析法，采用最近邻分析技术可以大体比较亲代与子代分子的一级结构，但不够准确。

如果，想象一条由许多 A、T、G 和 C 以任意次序组成的长链，那么可以设想在此 DNA 分子中鸟嘌呤与腺嘌呤邻接的机会是有限定的，即有一定的邻接频率，同样，腺嘌呤也会在 DNA 分子中以一定的频率与胞嘧啶相接，照此方式，四种碱基总共可形成十六种邻接二核苷酸组合，在完整地分析一条单链 DNA 之后将可获得一系列有关邻接二核苷酸对组成的频率（如GA，GT，GC的频率，AC，AG，AT 的频率如此等等）。最近邻分析技术就是一种测定 DNA 分子中每个二核苷酸对频率的方法。

现将这一技术的操作程序简述如下。严格依照在 DNA 合成章节中所描写的方法，准备好四种反应混合物。每种混合物具有相同的 DNA 引物，四种三磷酸脱氧核苷以及Mg^{2+}等，在四种脱氧核苷酸中，只有一种用同位素标记，各种不同的混合物所标记的脱氧核苷酸各不相同。例如，混合物 1 标记的是 dATP，混合物 2 标记的是 dCTP，以此类推。待四种混合物反应完毕，取得 DNA 鉴定

后发现在新合成的每一种 DNA 中^{32}P系位于那个特定标记的脱氧核苷(Y)和与之最近邻的一个核苷(X)之间(图 64)。^{32}P 虽然位于放射性脱氧核糖核苷酸和与它邻接的核苷酸之间，那么，现在该如何去鉴定那个与之相邻接核苷酸(X)呢？幸而找到了一些酶，例如小球菌 DNA 酶和磷酸二酯酶能够专一性地切断脱氧核糖核苷酸之间的糖-磷酸键，以使 DNA 分子降解。从化学反应来看，脱氧核糖核酸酶催化的水解反应乃是使原先和标记核苷酸(Y)连接的^{32}P磷酸分子转而与其邻核苷酸(X)的 3′位相连。水解以后，鉴定放射性核苷酸就能够一目了然查明与原来带有^{32}P的核苷酸(Y)相连接的那个核苷酸(X)。混合物 1 中的Y核苷酸为腺嘌呤核苷酸，混合物 2 中的Y核苷酸为胞嘧啶核苷酸等等(图 64)。

分析了最近邻核苷酸技术设计得以成功的有效方法和严密推理的情况，人们将能据以获得关于 DNA 分子结构的大量可信的

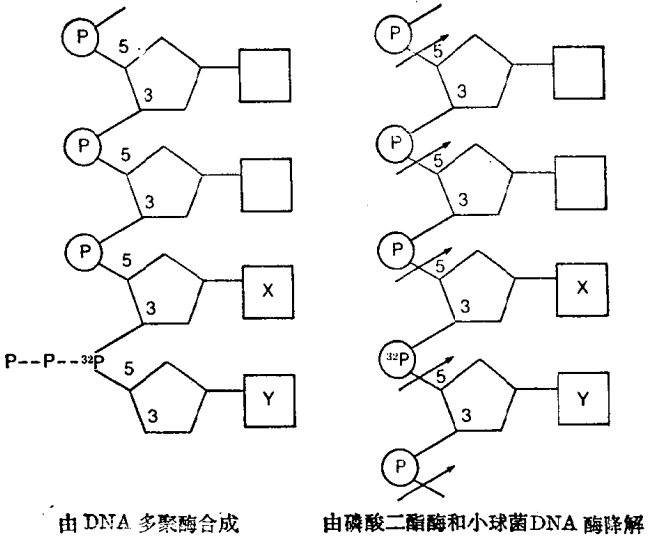
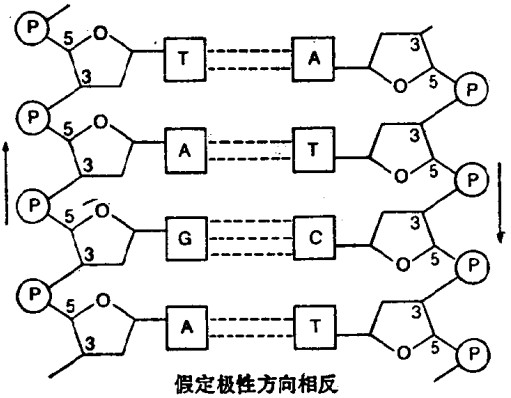


图 64　鉴定 DNA 分子中核苷酸顺序的最近邻分析方法

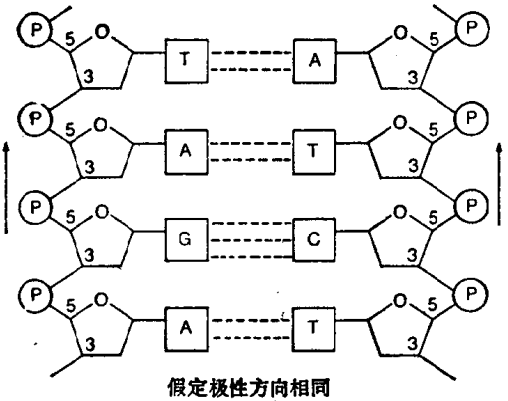
图中可见，^{32}P标记已经组合进DNA分子位于三磷酸脱氧核苷酸Y与X核苷酸衔接处，当DNA被适当的专一性酶于图中箭头所指处切断之后，^{32}P 即转附于与 Y 最邻近的核苷酸 X 的 3′ 位置上。(引自 J. Josse, A. D Kaiser 与 A. Kornberg, J Biol.Chem. 236:864, 1961。)

资料，从而对 DNA 双螺旋模型在认识上又取得新的进展(图 65)。在 DNA 双螺旋分子中，两条链的极性取向相反，亦即在双螺旋的一端，一条链为 3′末端，另一条链为 5′末端。不论在双链中那条链



假定极性方向相反

TpA (0.012) = TpA (0.012)
ApG (0.045) = CpT (0.045)
GpA (0.065) = TpC (0.061)



假定极性方向相同

TpA (0.012) = ApT (0.031)
ApG (0.045) = TpC (0.061)
GpA (0.065) = CpT (0.045)

图 65 DNA 分子中的两条链的极性

应用最近邻方法可以获得DNA分子中二核苷酸对频率的数值，假定 DNA 两条链的极性方向相同和假定极性相反，所估计的频率是不一样的。如果按照二条链的极性方向相反来计算，则观察值（左边括号内的数据）和预期值十分接近，而按照两条链的极性方向相同，即取向平行，则观察值与预期值有很大出入。(引自 J. Josse, A. D. Kaiser, 与 A. Kornberg, J. Biol. Chem. 236:864, 1961)

上，对每一个脱氧核苷酸来说都有一个3′近邻核苷酸和一个5′近邻核苷酸。新链的合成总是从5′到3′进行的，这暗示着两条链的合成方向彼此相反。近邻核苷酸系指3′位核苷酸而言，底物脱氧核苷酸(Y)的放射性^{32}P标记在5′位上，在底物分子结合进生长着的DNA链之后，便使^{32}P与近邻核苷酸(X)3′位相连。

每类起反应的混合物可以产生四种频率互不相同的二核苷酸对。四类混合物总共产生有十六种二核苷酸对的组合方式(图66)。Kornberg就核苷酸对频率在不同组合有两两相似，例如GpA=TpC，TpA=ApT等等的情况，推断亲代DNA分子是作为模板，而不是作为引物起作用的。简言之，模板分子的信息，在DNA合成期间，通过转录相同的碱基顺序，而转移给子代的DNA分子。然而应该指出，模板DNA单链与其DNA子链的碱基顺序并不相同，但表现为彼此互补。由此可见，当亲代DNA分子中互补的双链在DNA合成时，各自作为二条新链的模板，其所产生的新的子

	A	G	T	C
A	ApA 0.024	ApG 0.045	ApT 0.013	ApC 0.064
G	GpA 0.065	GpG 0.090	GpT 0.060	GpC 0.122
T	TpA 0.012	TpG 0.063	TpT 0.026	TpC 0.061
C	CpA 0.063	CpG 0.129	CpT 0.045	CpC 0.090

图 66　子代 DNA 分子中二核苷酸对的频率

图中所列数据为使用近邻分析法所设计的四类混合物反应中从子代分子所测的十六种二核苷酸频率。例如在 *Mycobacterium phlei* 的DNA中腺嘌呤与腺嘌呤邻接的频率为0.024，而鸟嘌呤与腺嘌呤邻接频率是0.065，如此等等。如DNA双链的极性方向相反，则合成方向亦相反，因新链合成方向为5′→3′。据此推测，应当GpA=TpC，TpA=ApT，ApA=TpT等，从实际得到频率来看，结果与预期的完全相符。(引自 A. D. Kaiser, 与 A. Kornberg, J. Biol. Chem. 236:864, 1961。)

代 DNA 双螺旋分子必然会与亲代 DNA 双螺旋如出一辙。

由最近邻分析实验可以看到，DNA 双链分子中的两条单链极性相反和顺序互补乃是确实可靠的事实。例如，可以预期，DNA 分子中近邻二核苷酸对的频率，在假定双链走向相同（即极性相同）的情况下，与假定双链走向相反的情况下是不一样的（图 65）。只有按双链极性相反计算，由氢键对接的二核苷酸对的频率才会一致，亦即TG的频率等于 CA 的频率。如果按双链极性相同计算，则TG的频率必然不等于 CA 的频率。〔从实验结果来看，如果双链的极性相同，例如 $\downarrow\begin{matrix}T\\G\end{matrix}-\begin{matrix}A\\C\end{matrix}\downarrow$，则 TpG 形成的频率为 0.063，ApC 形成的频率为 0.064，如果双链的极性相反，例如 $\downarrow\begin{matrix}T\\G\end{matrix}-\begin{matrix}A\\C\end{matrix}\uparrow$，则 TpG 为 0.063，CpA 为 0.063，可见，同样的二核苷酸对，其形成的频率是依双链中二条单链的走向而异的。也就是说，只有双螺旋单链极性相反才能使T与G邻接的频率等于C与A邻接的频率，即 TpG(0.063)＝CpA(0.063)。如果极性相同，这二种二核苷酸对形成的频率便不相等。以上只是举的二核苷酸对 C与A的例子。——译注〕其实任何一种二核苷酸对都同样如此。因为对某一个特定的二核苷酸对来说，都是遵循 A-T，G-C 的配对原则。〔这样，与之相对的那个位于互补链上的二核苷酸对必然是一定的（如 GA 来说，互补链上的二核苷酸便一定是 CT 等等）。两者形成的频率理论上也应该相同。——译注〕

并非所有有关 DNA 合成的问题都由于揭示了 DNA 子链与亲链分子碱基顺序的相似性而水落石出（表 5），事实上，在检测子链物理性质的过程中，又发现了许多新的疑难。例如，在比较亲代 DNA 分子与子代 DNA 分子的沉降特性时，即可发现子链的分子量在 4×10^6至 6×10^6道尔顿之间，而亲代 DNA 分子的分子量却为 8×10^6道尔顿。此外还有，在一般情况下，75℃左右的温度能使 DNA 分子变性，此时双螺旋分子的二条单链即彼此分开，再缓慢冷却，经变性处理过的 DNA 单链的混合物又会使二条解旋的互

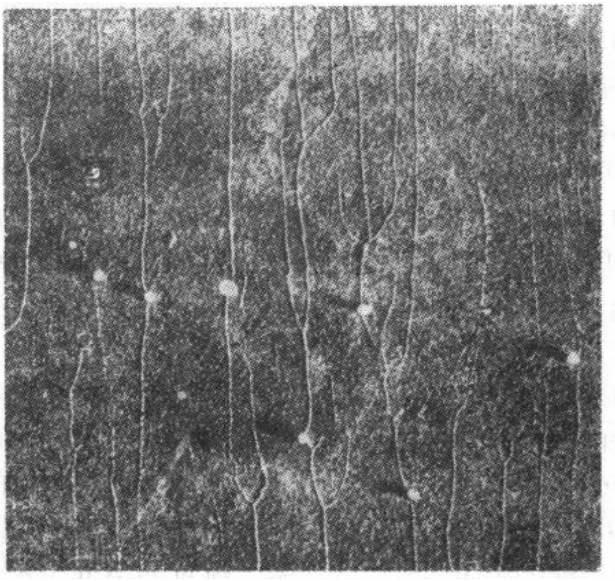
补单链重新恢复双螺旋结构，这便是所谓的互补单链的退火。在这里如果将亲代 DNA 分子与子代 DNA 分子的退火性作一番比较，便会发现子代DNA 分子的退火速度较之亲代 DNA 分子退火的速度快捷得多。现在要问，如果子代DNA分子肯定是亲代DNA分子的准确复制品，那么二者为什么会在物理性质上有如此的差别呢？为了寻求这个问题的答案，对亲代 DNA 分子及子代 DNA 分子进行了电子显微镜的观察。在电子显微镜下，可以看到亲代 DNA 分子是直链形的分子，而子代 DNA 分子则为支链形分子(图 67)。

有人在评论此事时提出了疑问：既然子代 DNA 分子与亲代 DNA 分子不尽相同，那么体内 DNA 合成实际上是否果真按

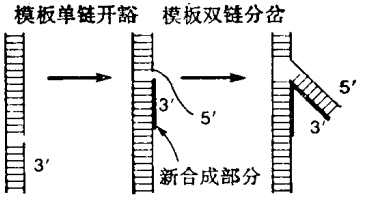
表 5　使用不同种属的 DNA 作为模板，通过酶促反应合成的 DNA 分子的化学成分*

DNA	A	T	G	C	$\frac{A+G}{T+C}$	$\frac{A+T}{G+C}$
Mycobacterium phlei:						
模板	0.65	0.66	1.35	1.34	1.01	0.49
产物	0.66	0.65	1.34	1.37	0.99	0.48
大肠杆菌:						
模板	1.00	0.97	0.98	1.05	0.98	0.97
产物	1.04	1.00	0.97	0.98	1.01	1.02
小牛胸腺:						
模板	1.14	1.05	0.90	0.85	1.05	1.25
产物	1.12	1.08	0.85	0.85	1.02	1.29
T_2 噬菌体:						
模板	1.31	1.32	0.67	0.70	0.98	1.95
产物	1.33	1.29	0.69	0.70	1.02	1.90

* 引自 Kornberg, 1974。



(a)



(b)

图 67　电子显微镜下的子代 DNA 分子

(a) 由大肠杆菌多聚酶合成的子代 DNA 分子是支链形分子，而不是直链形分子。(本图片由 Arthur Kornberg提供，电镜摄影 Ross Inman。)

(b) 图中模型表示支链产生的一种可能机制。

Kornberg 所设计的系统那样进行呢？考虑到在此系统中所形成的子代 DNA 的二核苷酸对频率确实与亲代DNA分子如出一辙，不容置疑，因此，上面提出的问题，不免令人感到惊异。其实这个问题的真意是问：新合成的DNA是否也与体内合成一样，是一种具有功能的DNA，亦即它是否也同样具有生物学活性呢？这个问题

又进而引伸出第二个问题：能够可靠地测定新合成 DNA 生物活性的试验，应该具备哪些条件？

当时能够令人信服地检验 DNA 活性主要指标之一，显然是其转化活性，有人设想利用已知的具有转化活性的 DNA，例如肺炎球菌 DNA 作为模板，人工合成子代 DNA 分子，然后再比较分析子链和亲代 DNA 分子的转化活性。但是，这一旨在合成有转化活性 DNA 的类似图谋大多数都失败了，在所有的实验中，所合成的子代 DNA 分子在电子显微镜下都毫无例外地一律呈现支链形状。

在与 Kornberg 就新合成的 DNA 寻找某种行之有效的功能测验方法的同时，Robert Sinsheimer 研究了一种很小的噬菌体 ϕX174。ϕX174 之所以引人注目，这是因为该噬菌体染色体为一条单链 DNA 分子，属于已知最小的染色体之列。其长度为 6000 个核苷酸，按基因的一般大小计数，约可包含 6 至 8 个基因，ϕX174 的另一个引人入胜的特点是它能以“裸露”的 DNA，也就是能够不以整个的噬菌体颗粒感染大肠杆菌成功。这些特性能使 Kornberg 领悟到 ϕX174 DNA 不仅能为试验工作提供便利，而且也是可作为设计检测 DNA 的生物活性的可靠材料。作者以 ϕX174 作模板的系统中，结果终于合成了能够感染大肠杆菌成功的新的 DNA 分子(图 68)。在此有必要特别指出，ϕX174 DNA 本身是一种环状染色体，而不是线状染色体，而新合成的 DNA 却仅为线状分子，但是，线形的子代 DNA 分子并不具备染色体的功能，线状分子在催化线状分子环化的 DNA 连接酶的作用下，可以变为环状分子。只有环状的 DNA 分子才能够成为有功能的染色体。

然而，具有生物活性的 ϕX174 DNA 的合成成功，并未解决全部疑难。例如，为什么迄今不能合成有转化功能的 DNA〔1977 年 11 月份报道，博耶等人已人工合成一个称为生长激素抑制因子(Somatostain)的基因，能够通过质粒转化和表达——译注〕的原因，仍然不明。此外，还发现 DNA 分子在体内和体外的合成速率

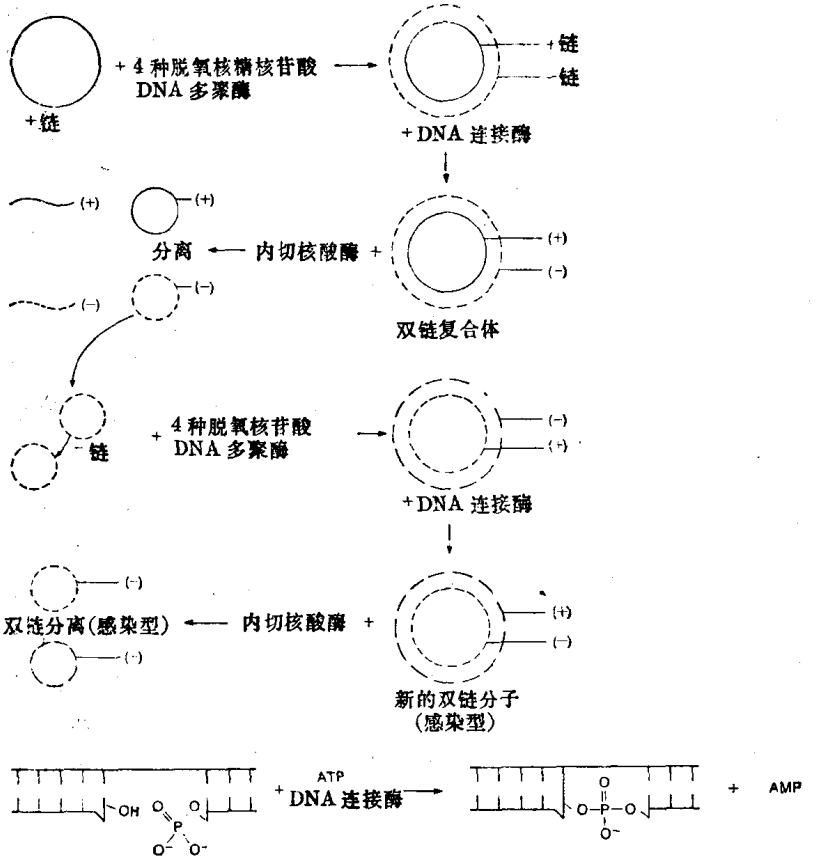


图 68 由酶促合成具有生物活性的噬菌体 DNA

φX174新链 DNA 分子合成步骤，开始时先从 φX174 噬菌体颗粒分离出环状单链，称为(+)链，合成结束产生出许多有活性的(+)链，图示从原先的(+)链合成(−)链，多核苷酸连接酶再使(−)链环化，产生出环状的双链分子，在双链的环状分子中加入内切核酸酶，又切开双链中的(+)链或(−)链，由于(−)链中的胸腺嘧啶被5-Bu(5-溴尿嘧啶)替代，故(−)链较重，密度梯度离心法可将两条链分开，此法处理一批双链分子，可以获得四种不同类型的分子：完整的环状(+)链和(−)链，线状的(+)链和(−)链分子。然后，再以环状(−)链作为模板合成新的双链分子(在多聚酶的作用下)再将第二次合成的双链分子用内切核酸酶切开。通过密度梯度离心分离(−)链，用连接酶将其环化，即可作为具有感染能力的 DNA，图中底行说明的是连接酶所催化的环化反应的细节。(引自Kornberg和 Goulian。)

大不一样，在大肠杆菌细胞中每分钟的合成速率接近于 90,000 个碱基对，或相当于长度为 3μm 的 DNA 分链。可是象这样的合成速率却并非是任何细胞都能赋有的。譬如，仓鼠细胞 DNA 分子合成的速率只有每分钟 2.5μm，但即使如此，仍然为离体系统时所望尘莫及。后者的合成速率平均每分钟约为 1000 个碱基对左右，DNA 在合成速率上所表现的差异，使人们注意到以下可能性：要么是催化离体系统反应的酶并不是活体内催化同一 DNA 分子合成反应的酶；要么就是用于离体合成试验的条件在一个或多个方面很可能远远次于活体内所达到的最佳状态。

另一个令人困扰的问题渊源于合成有活性的 ϕX174 染色体的多聚酶。为了说明这一点，有必要考察一下 DNA 分子复制的过程(图 69)。在图中可以看到同时进行的有两方面的过程：一个是亲代 DNA 分子的解旋伸展，另一个是在模板引导下半保留合成两条新链，每条新链各自与一条旧链进行互补。亲代 DNA 环状分子在豁口相连处，亦即复制叉(交点)之前的部分，仍然由氢键互补，保持原有的双链构型，而在复制叉之后的部分，双链解旋拆离，作为模板合成新链。在两条分开的亲代 DNA 链上，同时进行着新 DNA 分子的复制合成，复制区段到交叉点为止，在合成期间，复制叉从一端向另一端移动，亦即从合成起始点向合成终止点移动，在两条亲本单链上，合成的区段亦随之齐头并进。

依据半保留复制的原则，复制叉前面部分全是亲代 DNA，而复制叉之后的两条 DNA 双链从组成上来看，其中均有一条单链为旧链和另一条单链为新链。这里产生了一个问题：既然亲代分子的极性相反，而 DNA 多聚酶又具有定向催化的特性(只能按 5′至 3′方向合成 DNA 分子)，新链一般是通过将 5′三磷酸核苷加到新链生长端的 3′位自由羟基上而向前延伸的；那么多聚酶究竟是怎样紧跟复制叉之后，在两条链上同时朝前(即方向相同)进行复制呢？(图 70 说明了利用电子显微镜去区分复制叉里面和复制叉外面的亲代分子与子链的情况。)

这些有待解决的问题促使着人们去进行思考和实验。在解释

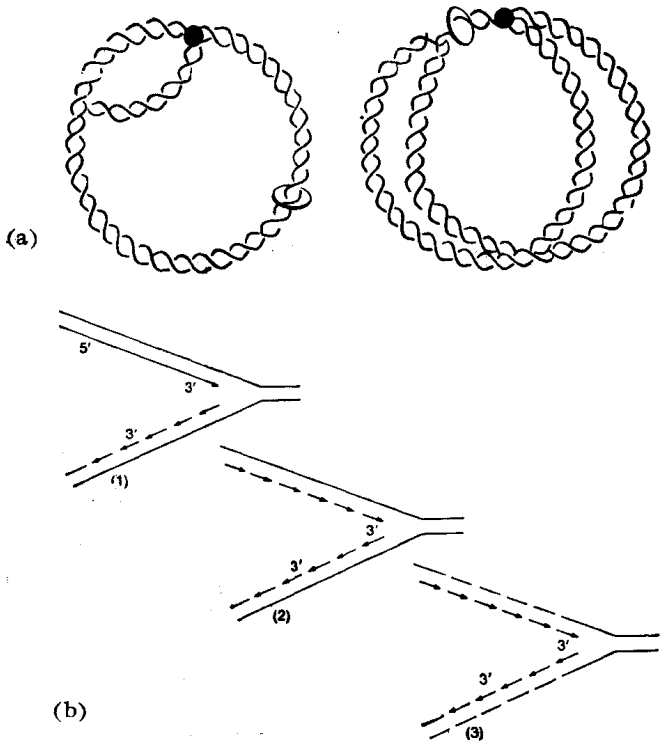


图 69　剖析复制期间的 DNA 分子

(a) 图示环状染色体复制过程中的二个阶段。(引自 Cairns, 1963。)

(b) 如果说合成只能从 5′至 3′，而且合成区段紧跟着复制叉齐头并进，那么复制叉之后的两股链又是如何同时合成的呢？是在一条链上进行不连续的合成呢(1)？还是在二条链上都进行着不连续的合成呢 (2)？还是在二条亲代链分子上切开许多切口同时进行合成的呢(3)？

上曾经提出三种可能的设想：(1) 可能是由于 Kornberg 使用的多聚酶并不是体内催化 DNA 合成的那种酶，(2) 虽然 Kornberg 使用的多聚酶按 5′至 3′方向进行合成，但细胞内可能另有一个合成方向为 3′至 5′方向的酶，亦即在体内合成中有可能不遵循实验中所发现的 5′至 3′的规则。(3)复制进程是不连续的，即多聚酶能够识别许多复制的起始点。在复制叉内二条臂上各个小片有一边是朝着复制叉的方向进行合成，另一边是背着复制叉的方向进行

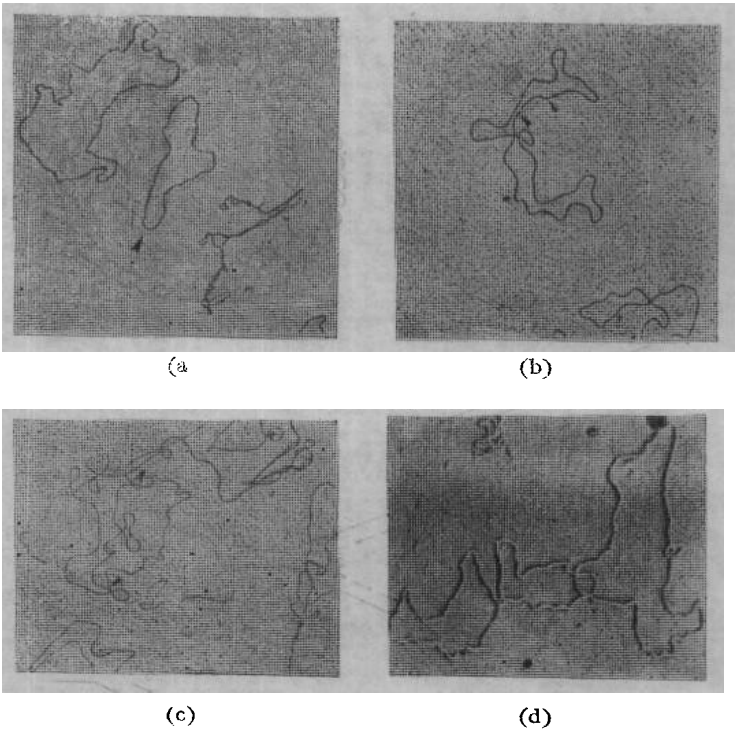


图 70　环状 DNA 分子的电子显微镜照片

(a) 图示果蝇(*Drosophila melanogaster*)线粒体中的闭环型DNA分子和开环型DNA 分子，DNA 总长约 6μm。箭头所指为标志的PM2 DNA，总长约 3μm。

(b) 鼠L-细胞线粒体DNA复制早期的中间体，二个箭头之间的区段为复制区域。

(c) 鼠L-细胞线粒体DNA复制晚期的中间体，二个箭头之间的区段是复制区域。

(d) 环连在一起的线粒体 DNA（环环相扣类似链索）。[(a)、(b)和(c) 由斯坦福大学 Clayton 赠送，(d) 引自 D. Clayton and C. A. Smith Int Rev of Exp Path 14:1-67, Academic, New York, 1975。]

合成，一旦在短的片段合成之后就通过 DNA 连接酶将之连接到子链上(图 71)。

在为揭开这些不解之谜而设计的实验中，曾研究了一种不能进行自身 DNA 复制的细菌突变型，并获得了最为圆满的结果，读者可能会感到惊奇：不能进行自身 DNA 复制的突变型通常是致

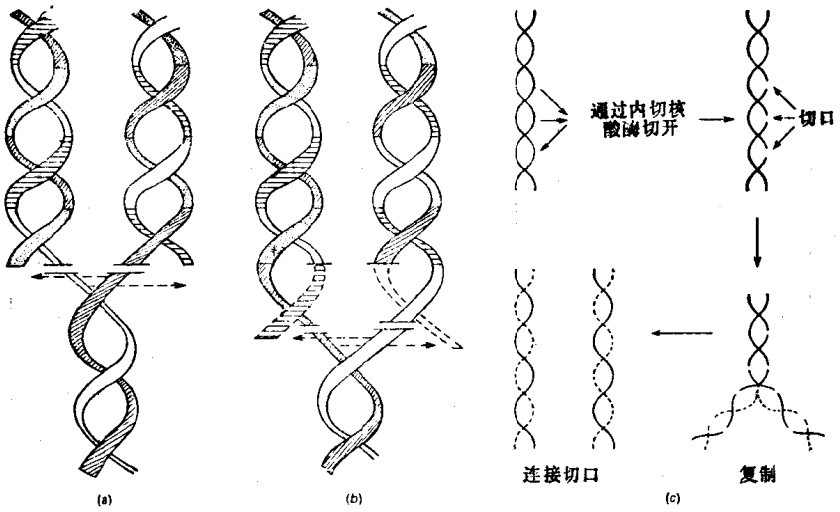


图 71 复制叉后面的 DNA 合成机制

1957 年 Max Delbrück 提出了一种假说旨在说明 DNA 复制期间如何通过主链的断裂和重接从而使双链得以分开的机制，以断裂重接方式进行的双链拆离，是从顶端开始自上而下完成的。DNA亲代部分(图中以白链和斜线链表示——译注)每解旋半圈便发生一次断裂，断裂形成的下面断端，与极性相同的子链(图中相应以点链和横线链表示)自由端连接，亦即白链与横线链相接，斜线链与点链相接。(a)和(b) 表示连接过程在时间上的二个阶段(相差半圈复制时间)，在这两个阶段之间亲代DNA解旋了半圈，也复制了半圈，虽然Delbrück当时并没有这一模板的实验证据，但是在设想上，它要比有些新发现要早。(c)这里包括内切核酸酶将链切开的作用机理，以及接下来由连接酶催化重接的作用机理等等。

死的，怎么可以用来作为研究材料呢？其之所以成为可能乃是由于许多突变型细菌在不同的环境条件有着不同的表型，有些不能复制 DNA 的突变型细菌在某种环境条件下招致死亡，但在另一种环境条件下却能存活。这样的突变型一般即称为条件致死突变型。细菌遗传学家曾设计了一些能用来筛选条件致死突变型的环境因子，其中最简单的环境因子便是温度。有一种条件致死突变型在 30℃温度条件下能正常生长，但在 40℃时即表现死亡，称为温度敏感突变型，这是在研究 DNA 合成的实验过程中经常使用的实验材料。

许多温度敏感突变型在 30℃时表现为野生型，但当温度超过 40℃时就不能合成 DNA。现已发现所有的突变型(只有一个突变型例外)一般都具有正常的 Kornberg 多聚酶，这意味着不能把 DNA 不合成的原因归结为缺少 Kornberg 多聚酶。然而，耐人寻味的是那个唯一例外的突变型其 Kornberg 多聚酶的活力，虽然只及野生型的 1%，在体内却能以正常的速率合成自身的DNA。通过进一步研究又发现该突变型的修复功能（修复遭紫外线损伤的 DNA 分子)已经完全丧失，由之暗示，该突变型所失去的酶乃是一种修复酶(已知大多数正常细胞均含有一定数量的修复酶，修复酶的作用是保证 DNA 分子在经诱变剂处理之后仍然能够完整如初)，确实，这个突变型对紫外光的敏感程度要超过任何其他的突变型和野生型，同时 DNA 分子链的重接能力也比较低。

对多聚酶缺失突变型所作的研究，使人们对 DNA 合成特征获得了新的认识，并由之判明了另外二种多聚酶的存在。对于目前一般习称为多聚酶 I(PolI)的 Kornberg 多聚酶的特性也有了更深入的了解。在这些探讨中取得的惊人成果之一是弄清了多聚酶 I 实际上有着多种多样功能。从其中某些功能来看，使若干过去无法说明的难题也能由之迎刃而解。已知多聚酶 I 的功能如下:

1. 合成质粒 DNA (图 72)。质粒 DNA 是在某些细菌中发现的游离于细菌染色体之外的小分子 DNA，能够独立复制。

2. 按照 5′至 3′方向修复某些遭到损伤的染色体 DNA。

3. 具有 3′至 5′的外切核酸酶的活性，即能在单链的 3′端每次切下一个未配对核苷酸。

4. 具有 5′至 3′外切核酸酶的活性，当有 5′末端碱基存在时(非闭合的 DNA 分子)，该酶才起作用(紫外线照射后，在 DNA 分子中会产生有胸腺嘧啶二聚体，多聚酶 I 的这一功能是在二聚体被切除的过程中发现的)(图 73)。

多聚酶 I 分子量大小约为 100,000 道尔顿，其中只有一个识别底物 DNA 的位置，在多聚酶 I 识别 DNA 之前，必须首先将分子切开一个裂口(双链中一股 DNA 单链的磷酸二酯键断裂)，或

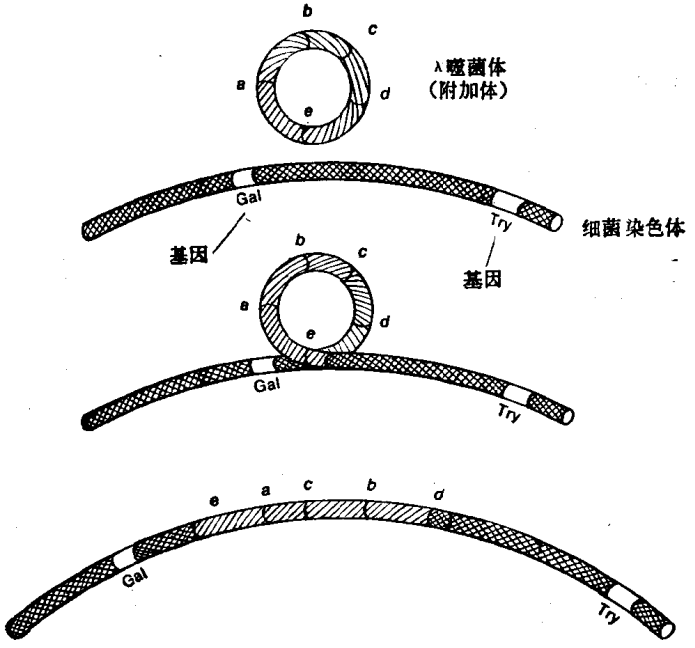


图 72 质粒 DNA

质粒DNA(又名附加体),大都为环状的小分子 DNA,质粒一般不依赖染色体DNA而独立存在,偶尔也会整合到染色体DNA上,图中所示为λ噬菌体DNA的特征:既可以游离于大肠杆菌染色体之外,也可以整合在大肠杆菌染色体之上。这是Allan Campbell 提出的旨在解释质粒 DNA 整合到染色体上的一个模型。根据这个模型,质粒DNA可以和染色体DNA配对(联会),两者经过重组(交换)即可将质粒DNA插进细菌染色体。(引自 Stent, 1963。)

者是首先完成单链化。多聚酶 I 本身可以解离为二个不同的多肽链,其中一个多肽的分子量约为 76,000 道尔顿,另一个多肽的分子量约为 36,000 道尔顿。如果能在温和的水解条件下,将这二个多肽分开为二个独立的亚基(图 73),便会发现大亚基具有修复功能和有 3′至 5′的外切核酸的活性,小亚基则具有 5′至 3′外切核酸酶的活性。

现在已经公认,多聚酶 I 能够催化有功能特性的ϕX174 DNA的合成,因为多聚酶 I 可以"识别" ϕX174 的 DNA 分子并作为ϕX174 的修复酶而起作用(因为 ϕX174 为 DNA 单链,合乎修复

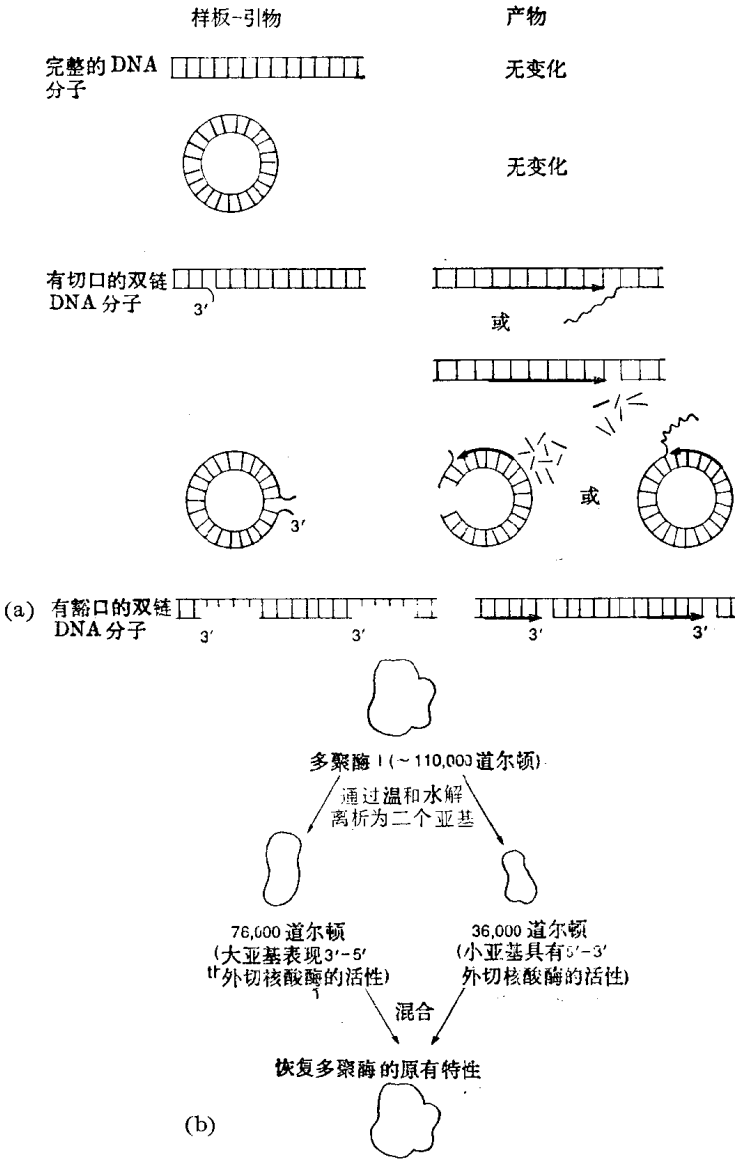


图 73　多聚酶 I 的催化性质

(a) 多聚酶 I (Kornberg 多聚酶)所表现的五种催化功能。可参阅正文中有关多聚酶 I 活性说明。(引自 Kornberg, 1974。)

(b) 在温和的蛋白质水解条件下（指使用蛋白质降解酶进行部分消化）可将多聚酶 I 离解为分子量为 76,000 道尔顿和 36,000 道尔顿的二个多肽，每个亚基都具有不同的催化特性，二者一旦混合，又可恢复原先的催化功能。

酶的作用条件，而单链修复的结果便是复制——译注)。ϕX 174 的 DNA 又只有一个识别位点。这样在修复时总是从一点起始，不致于从几点起始，产生不完整的复制品。多聚酶 I 一旦附着到识别位置上，便可以从起始端到 3′ 末端一气呵成地完成整个 DNA 链的修补合成(复制)。子代 DNA 复制完成之后，剩下的使分子活化的最后一道工序，就是由 DNA 连接酶将线形 DNA 分子的两端连结环化，最终成为一个完整的环状 DNA 分子〔至此合成的为(－)链。(－)链与(＋)链一般都具有可感染寄主的生物活性——译注〕。

虽然有关 DNA 合成的早期实验实际使用的是修复酶（多聚酶 I)而不是"真正"的聚合酶（现在知道这是多聚酶Ⅲ)。但是有关 DNA 的大量资料，诸如 DNA 的结构、修补合成的机制、DNA 的各种类型特征等等，却都是借助于它，这件事又一次生动不过地证明了为什么科学的进步不可能仰赖于准确无误的安排，在更大程度上它所需要的更毋宁说是持续不断的尝试与失误。有时候，某种错误的概念甚至在一段很长时期会为人们所接受，从这一意义来看，由于我们有关 DNA 合成的知识都是根据离体实验的结果，这一结果是否真正能说明体内 DNA 如何合成的确切细节也很难说，恐怕还得有待于今后的实践。

图 74 概括地表示了目前所了解的 DNA 合成的基本模式。合成之前，内切核酸酶在亲代 DNA 分子的一股链上造成一个切口，使 DNA 模板暴露出 3′-OH 游离基。然后，多聚酶 I，即修复酶从这一切口开始通过合成与另一股亲本 DNA 分子互补的一小段新 DNA，跻身于原先亲代互补链的位置，随之，连接酶又再将亲代 DNA 分子链的断头重新接好。于是在岔口的后面便形成了一段由四股单链组成的中间体，其中有二股新合成的 DNA 分子链，一旦新链由连接酶连接完毕，复制叉即便脱颖而出，复制叉的双臂全部为双螺旋结构。

DNA 复制的全过程看来比较复杂，但是图中描绘的步骤与实际观察到有关的多聚酶及内切核酸酶的情况还是一致的。在图 73

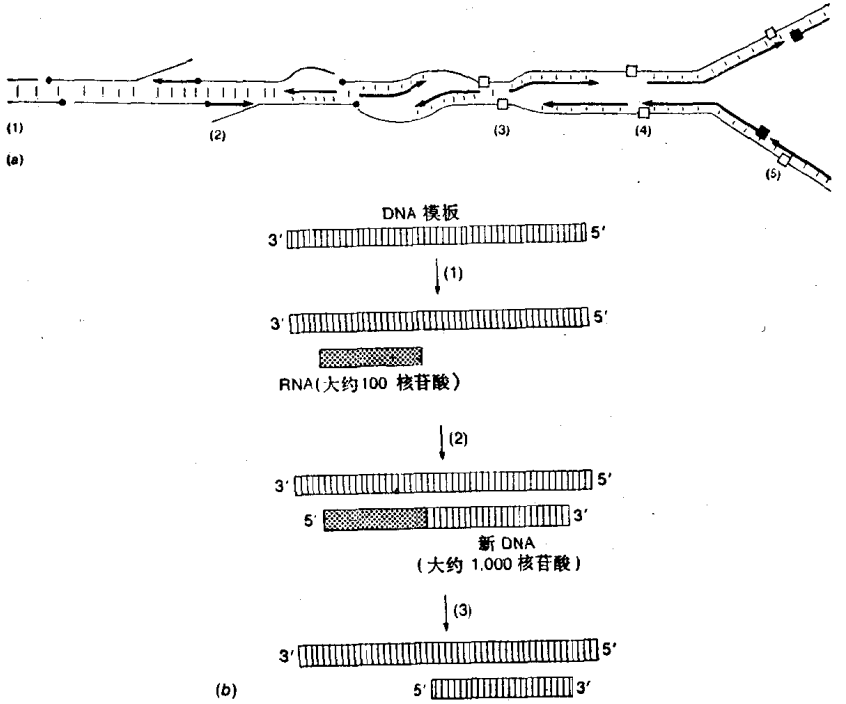


图 74 DNA 体外合成示意图

(a) DNA体外合成包括下列步骤: (1)内切核酸酶使单股链断裂产生切口，使DNA模板暴露出3′-OH 游离基; (2)由多聚酶 I 修复合成产生DNA片段替代原先的亲代DNA; (3)连接酶将亲代DNA的断头重新接好; (4)形成四链式DNA的中间体; (5)新的DNA片段由连接酶相互连接: ●＝内切核酸酶在亲代DNA上的切口，□＝亲代DNA切口重新相连，与原先位置相应，■＝新DNA单链的复制起始点，→为新DNA单链，——为亲代 DNA。(引自 Haskell 与 Davern, 1969。)

(b) 要有RNA作为必要的引子或启动因素才能引发上述DNA 复制步骤得以进行。(1)RNA多聚酶合成一小段RNA(约 100 个核苷酸)，这小段RNA与DNA的模板的一条单链的一小段互补; (2)RNA末端核糖核苷酸的 3′-OH是多聚酶Ⅲ合成新DNA(约 1000 个核苷酸)的片断的引物; (3) 随着DNA-RNA链的RNA部分水解，最后只剩下新合成的DNA分子。

中曾经罗列了有关多聚酶 I 的各种活性。这里还有第二个修复多聚酶，即 PolⅡ(多聚酶Ⅱ)，它能够以 5′至 3′方向修复 DNA 分子中较长的片段，但是此酶无外切核酸酶活性。“真正的”复制酶实为多聚酶Ⅲ(PolⅢ)，多聚酶Ⅲ合成 DNA 分子的方向也是从5′至

3′。不过对于多聚酶Ⅲ的性质几乎一无所知。正常的细菌细胞多聚酶Ⅲ的含量往往不超过 10 个分子，所以研究多聚酶Ⅲ极为困难，现将目前仅知的有关此酶的点滴性质列入表 6 及表 7。图 75 展示的亦只是一个大体轮廓。

表 6　大肠杆菌多聚酶Ⅰ、Ⅱ、Ⅲ的性质*

	PolⅠ	PolⅡ	PolⅢ
功能:			
5′→3′聚合作用	+	+	+
3′→5′外切核酸酶活性	+	+	+
5′→3′外切核酸酶活性	+	−	−
对各种模板-引物表现的活性反应:			
完整的双链分子	−	−	−
作为引物加入的 DNA 单链分子**	+	−	−
有切口的双链分子[采用多核苷酸d(A-T)]	+	−	−
双链 DNA 分子中间有单链豁口或拖有单链末端:			
单链部分从 5′末端算起核苷酸数目在 100 以下	+	+	+
单链部分从 5′末端算起核苷酸数目在 100 以上	+	−	−
完全不需要引物(重新合成)	+	−	−
一般性质:			
分子量(道尔顿)	109,000	120,000	180,000
每个细胞可能存在的分子数目	400	100	10
是否存在有条件致死突变型	有	无	有

* 引自 Kornberg, 1974。

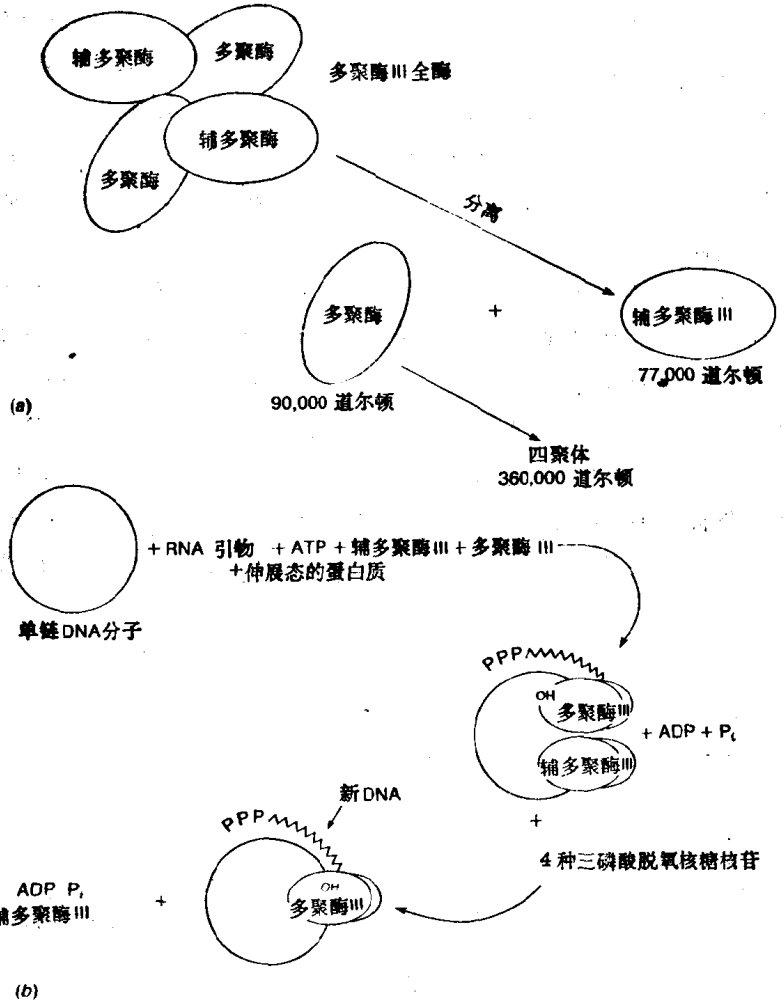
** 表示加入的单链是一条较长的单链，上面带有一小段在退火时形成的互补双链。

表 7　细菌中多聚酶 I 缺陷突变型的表型*

对紫外线敏感
对X射线敏感
修复遭紫外线损伤的 DNA 的能力极低
缺失频率增加
重组频率增加
φX174 在突变型寄主中复制时，复制型(双链形式)显示有单链缺口
突变型寄主缺乏 Pol1 外切核酸酶的活性，λ 噬菌体生长不良
T_4 噬菌体在突变型寄主内，丧失复制及重组能力

* 引自 Kornberg, 1974。

在大肠杆菌中，DNA 的复制过程似乎是不连续的。它先是合成一些片断，然后由 DNA 连接酶再把这些片断连接成子链。这样，产生的子链 DNA 分子是连续的、正规的，单核苷酸之间彼此由磷酸二脂键相连。不过关于这一复制过程的中间步骤迄今尚未完全建立，但可以肯定每个 DNA 片断 5′端上的 RNA 引物必须从中切除，这是其中的步骤之一〔这是由 DNA 多聚酶 I 行使外切核酸酶（从 5′→3′）的功能切除的〕。切除 RNA 后的空缺部分再由 DNA 多聚酶重新填补，至今所完成的动力学研究是支持这一看法的。用噬菌体 T_4 感染大肠杆菌作实验材料，加入 ^3H 胸腺嘧啶(即加入 ^3H 胸核苷)培养在短时间内(2 秒钟内——译注) 测定 T_4 DNA 的放射性脉冲标记，结果发现放射性仅仅只存在于约由 1,000 个核苷酸组成的较小的 DNA 片断上，这种片断以后即以其发现者命名，称为冈崎片断。如果过较长一段时间，再进行测定（30 秒钟以上——译注），可以看到分子量较大的 DNA 上也逐渐显示出放射性。在缺少 DNA 多聚酶 I（此缺陷型亦缺失 5′→3′外切核酸酶部分的活性，因之不能切除新合成 DNA 片断上的引物，影响片断的联结过程）或 DNA 连接酶的突变型中，冈崎片断的存留时间显著延长，从而能够使之充分显露出来，除此之外，通过其他多种突变型的研究也都表明其先复制小片断的过程完全相



(启动时 CopolⅢ 和 ATP 为必需物质，在 DNA 复制延续过程中则为非必需物质)

图 75 DNA 多聚酶Ⅲ的结构和功能

(a) 在DNA合成起始阶段，DNA 多聚酶Ⅲ是一个四聚体(即一个蛋白质分子由四条多肽组成)，多聚酶Ⅲ必须先与辅多聚酶Ⅲ(CopolⅢ)复合，才会发挥启动作用，在形成 PolⅢ-CopolⅢ 复合物后，便启动单链模板进行复制*。

(b) 合成开始时，整个混合物由环状双螺旋DNA，一个RNA引物，一个未折叠蛋白质，ATP以及 PolⅢ-CopolⅢ 复合物所组成，一旦合成开始，ATP 和 CopolⅢ 便不再是必需物质。(引自 Kornberg, 1974。)

* 据 A. Kornberg 所著«DNA Synthesis»一书论述，PolⅢ*加其他条件可以复制环状单链DNA。PolⅢ*与PolⅢ是状态上的差别，在一定条件下可以相互转化。本图文字说明为双螺旋 DNA，而图中为环状单链 DNA，以图中的环状单链较为确切。——译者注

同，除非作为中间产物的小片断的数量太少，无法检测，但即使如此，也还是可以从由之联成的分子量较大的 DNA 中去追查它们的存在。这一切都证明了冈崎片断实际上就是大分子 DNA 链的前体，一般来说，以放射性或其他方式"标记"的特定分子，对于它们以后的动向(如在转移、装配中的情况)也可在无标记的介质中加以追踪研究。

有关真核类生物细胞内 DNA 合成问题，虽然不准备在此细加讨论，但是，在高度进化的细胞内，DNA 的合成势必更为复杂，认识到这一点是十分重要的。例如，在真核类细胞的每条染色体所含有的启动部位就可能达 100 个之多，复制合成系朝两个方向展开，两个复制方向都各有一个复制叉 (图 76)。一般说来真核生物染色体的复制叉似乎与原核生物染色体复制叉大体相同，至于两者在化学结构水平上是否如此，则不太清楚。真核生物细胞的

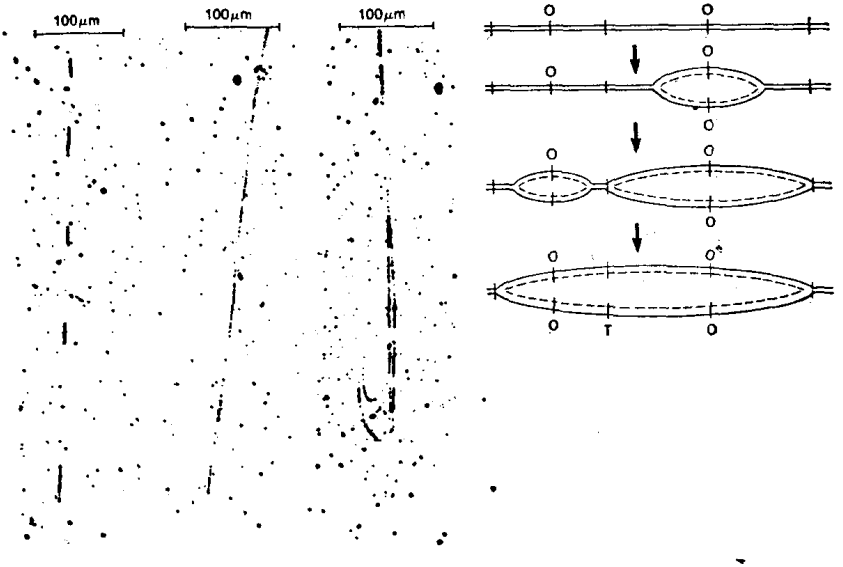


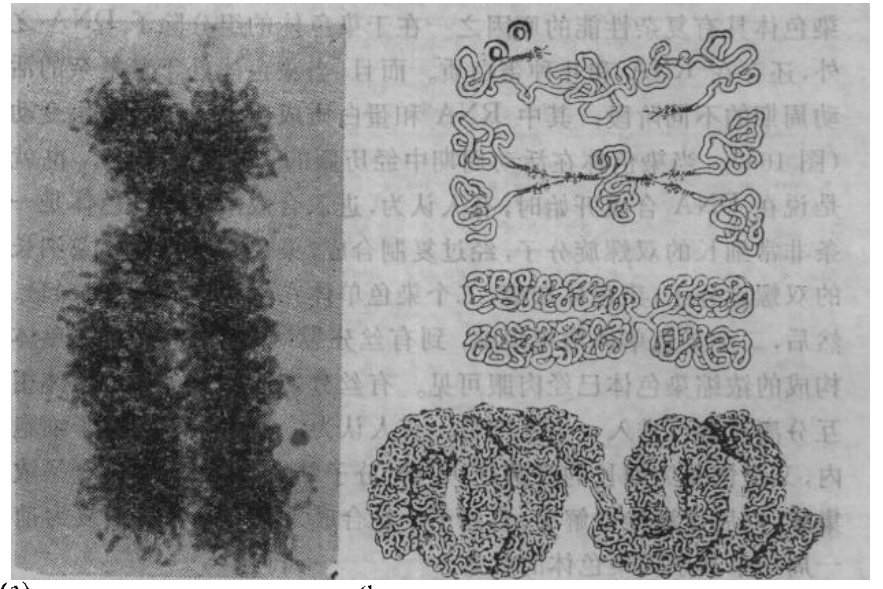
图 76　关于哺乳动物染色体的 DNA 合成

这是根据放射自显影试验结果提出的一张仓鼠染色体DNA合成的示意图，图中启动部位很多(图中为"O"处)，染色体DNA合成可朝二个方向展开，实验细胞用带有氚的胸腺嘧啶标记。(本图由 J. A. Huberman 与 A. D. Riggs 赠送，J. Mol Biol 32:327, 1968。)

染色体具有复杂性能的原因之一在于染色体的组分除了 DNA 之外，还存在 RNA 和几种蛋白质。而且，当染色体处于其复杂的活动周期的不同阶段，其中 RNA 和蛋白质成分也会随之发生变动(图 16a)。当染色体在活动周期中经历新的一轮变化之初，也就是说在 DNA 合成开始时，有人认为，进入合成阶段的染色体是一条非常细长的双螺旋分子，经过复制合成，染色体即成为二条细长的双螺旋分子(染色体加倍，二个染色单体在着丝粒区联在一起)。然后，二个染色单体增粗短缩，到有丝分裂期间，由二条染色单体构成的浓缩染色体已经肉眼可见。有丝分裂的中期，染色单体相互分离，各自进入一个子细胞。有人认为，染色体在新的子细胞内，又会恢复为细长的双螺旋 DNA 分子结构，起先折叠盘绕呈浓集状，以后逐渐伸展解旋，直到下一轮合成之前看起来又恢复为前一周期中其亲代染色体的状态。

染色体周史的整个细节还未全部知晓。事实上，周史中有几个阶段还有很多争议。例如，如果观察一下中期染色体的电子显微照片(电子显微镜图象)，就会看到，染色体似乎是由许多股细丝盘缠而成的一束纱线，在这里试图寻找染色体 DNA 链的线头——当然这谈何容易，但是很有可能会发现大多数染色体确实只有二个末端，这意味着每条染色体都是由一条细长的双螺旋分子自身来回折迭的产物，也正是因为多重折叠才在外表上看起来象是图 77 的模样。关于这一点也有很多不同意见，但正如在人类染色体的电子显微镜照片中看到的那样，这个问题的答案不是轻易能找到的。

Herbert Taylor 是研究真核生物染色体 DNA 分子合成的最杰出的科学家之一。作者通过自己建立的经典实验第一个为真核类生物的 DNA 半保留复制提供了证据，他以相当复杂的真核生物作为实验材料，实验方法如下：将蚕豆幼苗(*Vicia faba*)放在含有用氚作标记的胸腺嘧啶的溶液中培养（氚是氢原子的放射性同位素，胸腺嘧啶是合成 DNA 的前体之一)。用 ^{3}H 标记的胸腺嘧啶被幼苗根部吸收之后，最终会结合进幼根生长过程中分裂产



(a) (b)

图 77 DNA 在真核生物染色体内的包装方式

(a) 人类染色体的电子显微镜照片。
(b) 图示说明染色体中 DNA 分子包装的多种模型之一。(选自 E. J. DuPraw, DNA and Chromosomes, Holt, Rinehart and Winston New York, 1970。)

生的新细胞的染色体上。将植株在含有 ^{3}H 胸腺嘧啶的培养液中，经历几次细胞分裂，再将其转移到"冷的"(即无同位素)的培养基上，在这里随之出现的 DNA 合成只能摄取无放射性的胸腺嘧啶，使细胞在无放射性的胸腺嘧啶的培养液中只经历一轮细胞分裂，然后用放射自显影术检查根尖的染色体，用以观察 ^{3}H 的分布(图78)。由于 ^{3}H 发射的 α 粒子会在照相乳胶底片上产生黑点，因此即可直接显示出 ^{3}H 胸腺嘧啶结合进染色体上的位置。

在正常的根尖组织中，细胞分裂系在染色体复制之后，这意味着，由一次有丝分裂产生的子染色体最终分别进入两个子细胞内。因为染色体组进入到不同的细胞，判断染色体组之间的确切关系，很不容易。〔指在新形成的子细胞中要判定哪条子染色体是由哪条亲代染色体复制而来(无论是全保留复制或是半保留复制)是说

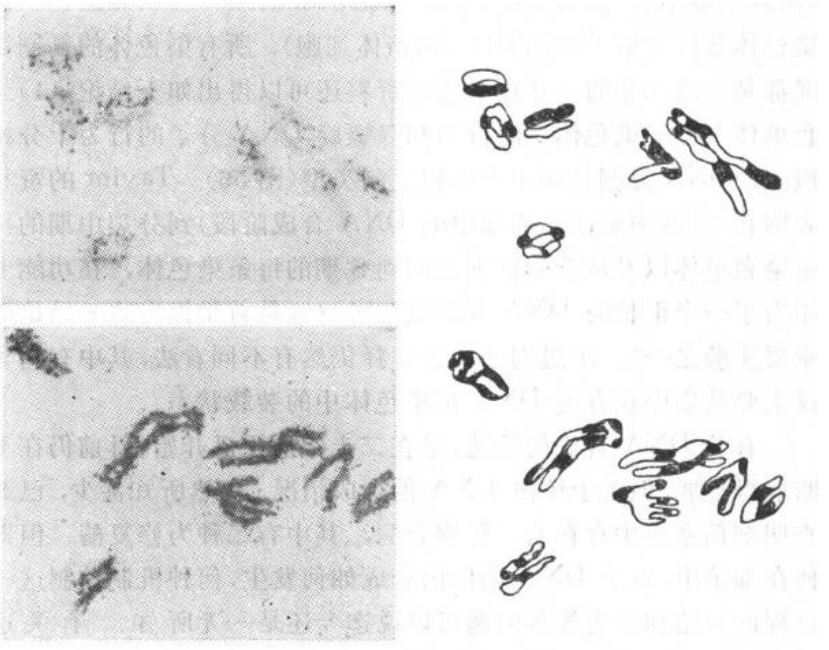


图 78　蚕豆根尖染色体的 DNA 复制

染色体上的黑点是由氚放射性粒子在乳胶片上曝光的产物。幼苗根尖在含^{3}H胸腺嘧啶的培养液中经过几次分裂之后，使所有染色体中的染色单体都带上了标记。然后转到"冷的"（无同位素），胸腺嘧啶培养液中，使细胞分裂（共经历 2 个细胞世代），可以看到染色单体的某些部分带有放射性，而另一些部分则不带放射性，从而为半保留复制提供了证据。右图为照片的解说性描绘。（由 J. H. Taylor提供，佛罗里达州立大学。）

不清楚的——译注。〕所以事实上当细胞经历几次分裂之后，很难追踪染色体本身的复制情况。为了解决这一难题，Taylor 在生长液中添加了秋水仙素。秋水仙素可阻止细胞进行有丝分裂（其作用是影响纺锤体正常形成），使染色体复制后无法拆离，进入不同子细胞。这样一来所有的染色体便都会存留在同一细胞之内，也就是说，秋水仙素具有一种特殊的能力，它能使细胞染色体数得以加倍。这意味着在一个细胞内所有的染色体其中二条染色单体的相互关系是一清二楚的，还可雄辨地说明，从起迄时间算起（即从标记溶液移到未标记溶液时算起），细胞中确实已经发生了一次

(而且不多不少,仅只一次)复制。如图 78 所示，可以确信在一个染色体数目加倍了的细胞内（四倍体细胞），所有染色体的复制情况都是一清二楚的。从这些实验资料还可以得出如下结论:(1)染色单体（即子染色体）的行为和双螺旋 DNA 分子的行为十分相似;(2)DNA 复制是属于半保留复制类型(图 56)。Taylor 的资料表明在S 期(细胞活动周期中的 DNA 合成阶段)到分裂中期的每条染色单体以及从分裂后期又回到S 期的每条染色体，在功能上相当于一个很长的 DNA 双螺旋分子（这是首批提出这一结论的重要实验之一)。不过对于这种解释仍然有不同看法,其中有些争议主要是集中在有关 DNA 在染色体中的装载状态。

有关 DNA 合成的论述,早在二十年前便已开始。目前仍在不断取得进展,但关于体内 DNA 的合成情况，仍然所知甚少，已经查明细菌系统中存在有三种聚合酶，其中有二种为修复酶。但即使在细菌中,对于 DNA 的体内合成如何发生,何种机制控制这一过程的起始和终止等等问题可以说迄今还是一无所知。〔有关这一问题的详细研究可参见 Kornberg 的《DNA 合成》(«DNA Synthesis»)一书。〕

突　变

如果从进化的角度来识别突变，可以说一切的原始材料都来自于突变，而自然选择则是自始至终指导生物演化的一种汰弱留强的作用过程。如果从孟德尔遗传学的观点来看待突变，则突变乃是使人们有可能在世代连续中观察基因表现的一种手段；亦即在孟德尔时期的那类实验中，每种基因至少都需要有二种相对的类型：一个为正常基因，一个为突变基因,否则便不能进行遗传分析。凡此种种,突变都是提供自然变异,即提供可遗传的变异。正是由于种内产生的自然变异,方使进化成为可能;也正是由于可遗传的变异,才会使血缘上相同的个体之间造成差异。因此,进化乃是一个使生物种属遗传性在长时期内发生的变化得以固定化的概念，而孟德尔遗传所研究的只仅仅涉及到有关个体之间一个或少

数几个基因的差别。

但是，就认识突变的理化机制或弄清等位基因之间在理化性质上的差别而言，并不一定先需要了解生物进化的过程和研究孟德尔式的遗传规律。而从理化性质方面来探明突变的机理亦就独自成了研究突变的第三个方面。虽然遗传学家多年来致力于突变机制的探索，但是在 DNA 的化学结构得以阐明之前，尚没有任何一个实验真正能够揭示突变机制的本质。

正如人们看到的那样，为什么肉眼可见的表型不能作为说明突变本质的理想指标呢？其基本理由是野生型基因和突变型基因都能够参与机体作用，使一系列代谢和生理反应倍加复杂，而这些反应又常常对整体表型的许多方面产生影响(图 79)。另外，非等位基因的不同突变也常常能够造成相似的表型变化（图 80)。这一看法之所以被人接受还在于这一事实：细胞和生物的发育所必需的代谢反应和生理过程都是在连续不断的环境因素作用下发生

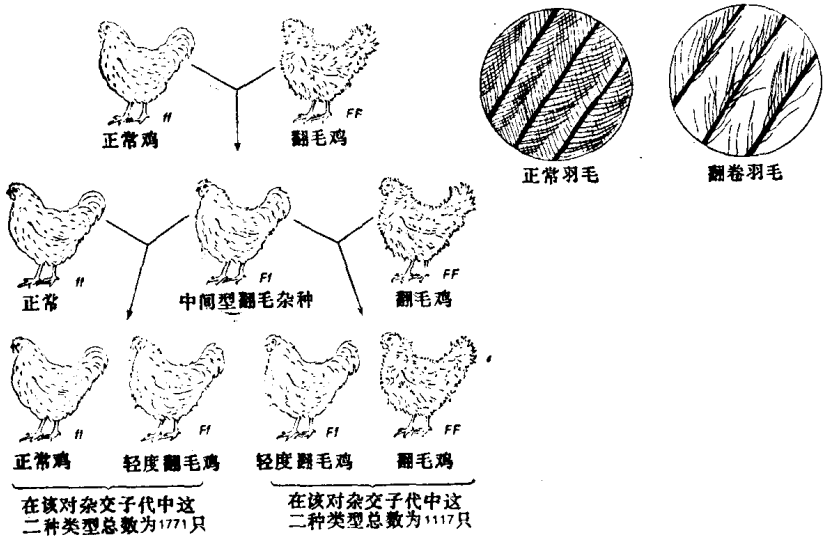


图 79　单基因突变可以产生多表型效应

图示一种家鸡单基因突变的遗传模型。该基因突变可产生一般称之为羽毛翻卷的表型。这一基因不但直接影响到羽毛的结构和质地，而且还具有多方面的效应：因翻毛鸡无法维持正常体温，因之表现心跳加速，心脏扩大，最后会引起多种解剖学性状的变异和生理学过程的异常。

图 80　非等位基因突变可造成相似的表型

人类皮肤的黑色素即为一例。黑色素的合成系按照一定的代谢顺序步骤，其中每一步均由不同的酶所催化。控制这些酶生成的基因是非等位基因。只要这些基因中的任何一个发生突变便都会产生白化症的表型(可参见图 19)。(野生世界图片。)

的，而环境因素的性质及其在作用上的时间顺序对机体表型的发育关系很大。

既然早期的遗传突变概念是在对类型研究的基础上提出来的，而不是根据突变本身的研究提出来的，因此毫不奇怪，那时关于突变的分类比较粗放，亦无内在机理作为依据。至于突变的分类系统则囊括了一切可通过表型改变而被认识的任何一部分的遗传物质变异和整个的遗传物质变异。从大的范围来讲，可以观察到整套染色体组的双倍化(多倍性)，整条染色体的增加或缺失(异倍性)以及染色体结构的改变等等(图 81 和 82)。从小的范围来看，一个突变可以发生在基因内部并能以单个的孟德尔因子的模式遗传，如镰刀形红细胞血红蛋白即为一例。

1953 年以后，对遗传突变的研究已进入到检测突变化学基础的阶段，即阐明基因内发生的能导致不同表型的有关化学变化过程。这方面最早的研究工作之一，是将某种化学结构上类似的物

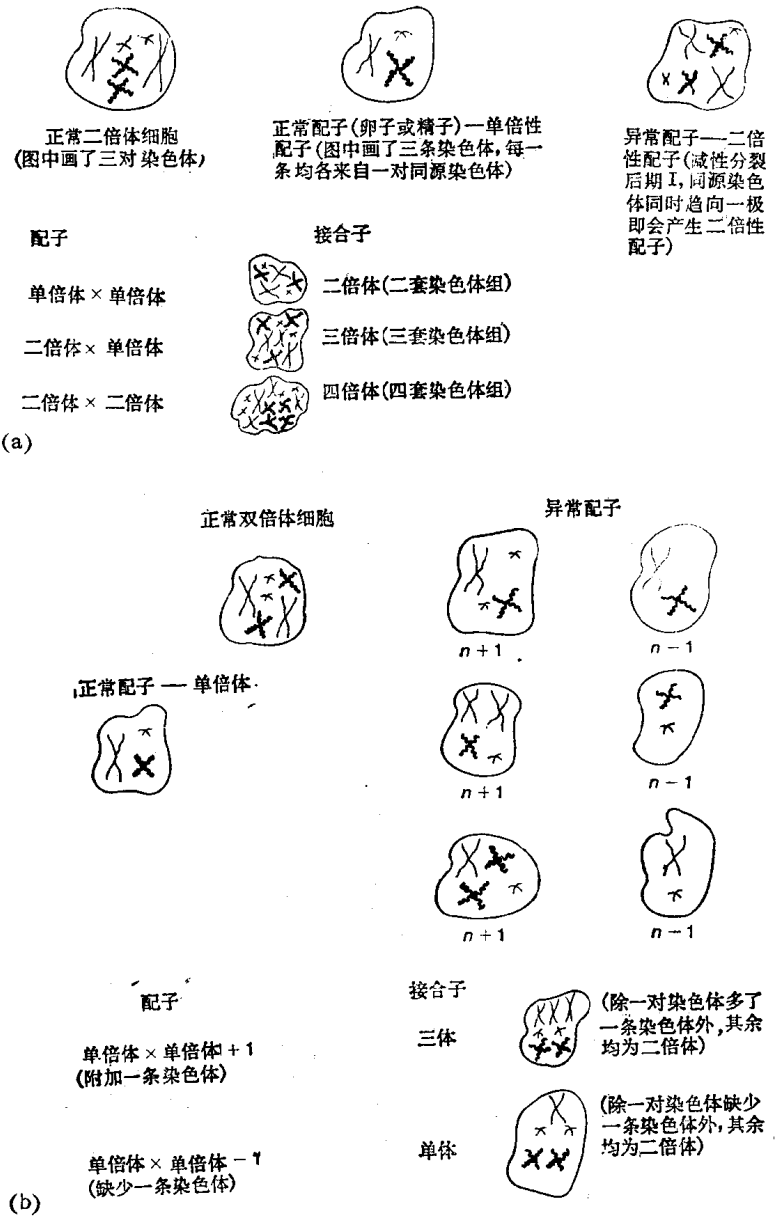
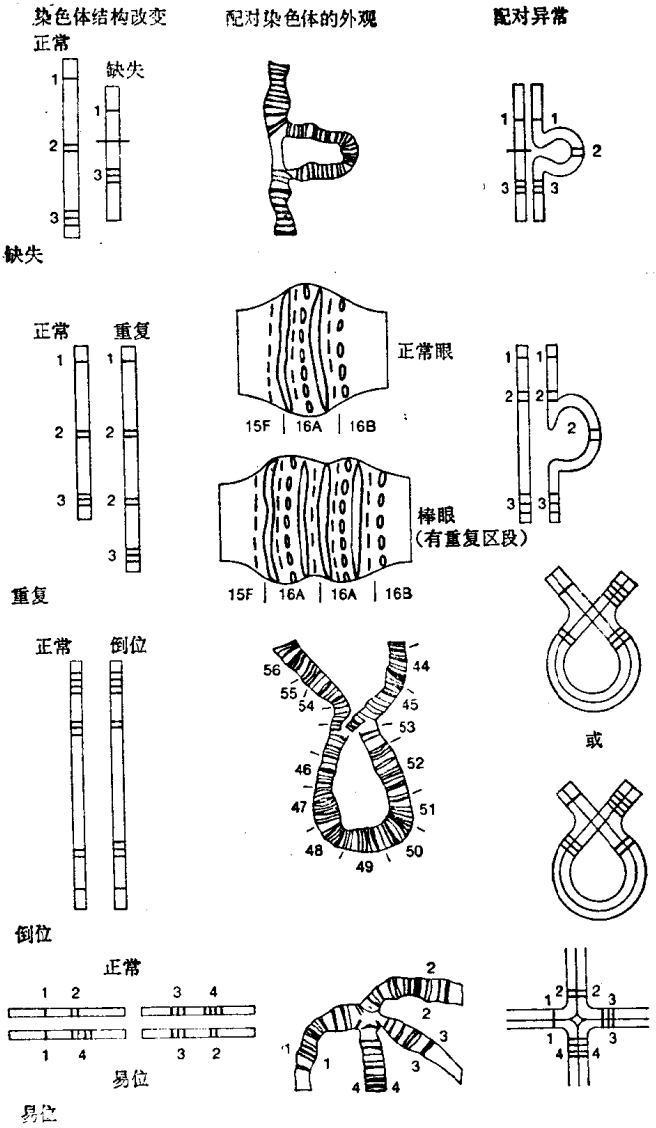


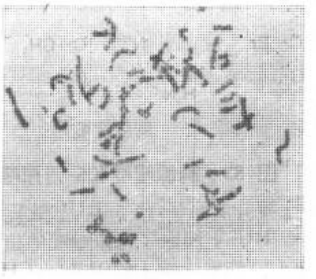
图 81 有关染色体数目变化的突变

(a) 多倍性。

(b) 异倍性。


染色体结构改变
配对染色体的外观
配对异常
正常
缺失
缺失
正常
重复
正常眼
15F | 16A | 16B
棒眼
(有重复区段)
15F | 16A | 16A | 16B
重复
正常
倒位
或
倒位
正常
易位
易位

(a)



(b)

图 82　有关染色体结构的突变

(a) 图中为四类基本的染色体结构变化。某些变异类型可以用光学显微镜进行观察，特别是在减数分裂早期当结构异常的染色体与结构正常的同源染色体进行配对时，图象尤为明显。

(b) 由化学因子诱导的人类染色体的结构变化。(由 T. C. Hsu, M. D. Anderson Hospital, Houston 赠送。)

质取代“天然”的嘌呤和嘧啶碱基使之结合进新合成的 DNA 链中。该项实验可产生在已知的化学组成上与正常 DNA 有所不同的新型 DNA 分子。

第一个被用来插入 DNA 分子的碱基类似物是 5-溴尿嘧啶(5-BU)。5-BU 的化学结构同天然的碱基胸腺嘧啶极为相似(图 83)。当 5-BU 加入供细菌细胞生长的培养基中时，即可结合进新合成的 DNA 分子中取代本应由胸腺嘧啶占据的位置。如果采用不能合成胸腺嘧啶的细菌突变型进行实验，则 5-BU 被结合的效率将会大大提高。因为这种突变型只能从外界获取全部胸腺嘧啶，故采用此突变型后，细菌所需的胸腺嘧啶基本上可由 5-BU 取而代之。

染色体的胸腺嘧啶位置已由 5-BU 取代了的细菌仍然可以存活和生长，这意味着置换本身并不能诱导突变；但另一方面，经这样培养的细胞，其每一基因都可能发生突变。事实上，已有证据表明 5-BU 在染色体内的功能很象胸腺嘧啶，它们之间有一个重要差别，即携带 5-BU 的 DNA 分子经证明比携带胸腺嘧啶的 DNA 分子有高得多的突变比例。尽管携有 5-BU DNA 发生的突变仍

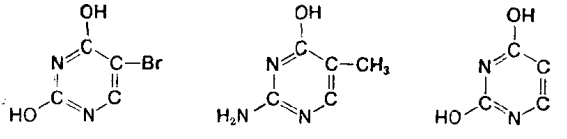


图 83 胸腺嘧啶和5-溴尿嘧啶

尿嘧啶的5位碳原子上接有溴原子，使5-BU较之尿嘧啶更象胸腺嘧啶；亦即，对DNA多聚酶来说，5-BU上溴原子与尿嘧啶的相同位置上氢原子相比看来更"象"胸腺嘧啶上的甲基($-CH_3$)。

然微乎其微，但从计算来看，它们要大大地超过正常的DNA分子发生突变的比例。

在5-BU的分子群体中含有二种形式的5-BU分子，或称互变异构体(图84)。一种为酮式分子，这是比较普遍存在的形式；另一种为烯醇式，一般比较稀少。但无论如何当5-BU作为底物取代胸腺嘧啶时，这二种形式的5-BU分子都会出现在新合成的

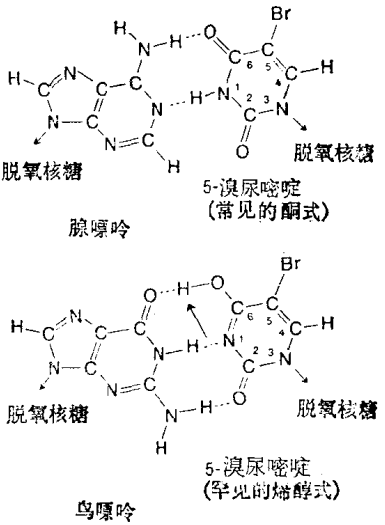


图 84 5-BU的两种构型(互变异构体)

在大量的5-BU分子群中通常有二种构型，或互变异构体。一种为酮式，数量上占优势(正常构型)；另一种为烯醇式，一般较为罕见。酮式分子能象胸腺嘧啶那样与腺嘌呤配对，而烯醇式分子类似于胞嘧啶，可与鸟嘌呤配对。

DNA 链中。数量稀少的烯醇式5-BU分子在 DNA 分子中，大概是定位在与鸟嘌呤相对的位置上（这是由氢键的结合特性所决定的），而不是象酮式分子那样与腺嘌呤对应。此种错配一旦发生之后，而且错配了的 DNA 分子又得以进一步复制，在存在胸腺嘧啶的场合，亲代分子的一条单链中原来与5-BU 配对的G此时仍然与碱基C配对，从而形成正常的 DNA 分子。而在另一条单链上由于5-BU 烯醇式异构体存在时间很短，很快又回复到一般常见的酮式异构体形式，而酮式的5-BU 链将与碱基A配对，新形成的互补链与亲代链相比，原先由G占据的位置此时已为碱基A所代替。待新链再次复制后，在第二代的互补链中，T 即插入与A相对应的位置，结果碱基对 AT 便代替了原来的碱基对 GC （图 85 a）。这种代换是基因突变的方式之一（参入错误）。

由一种嘧啶取代另一种嘧啶（例如由T取代C的情况）或者由一种嘌呤取代另一种嘌呤造成的突变通常称之为转换。发生转换的方向有 GC→AT 和 AT→GC 二种，但要用实验来区分它们比较困难。通过回复突变（即突变型中的突变又返回到野生型状态）的研究有可能决定第一次突变的方向性。例如，烯醇式5-BU 的参入错误可引起GC→AT 的转换，这样的突变，只有在5-BU 再一次引起 AT→GC 转换时，方才能为5-BU 所回复（图 85）。

如果将5-BU 诱导的突变再次用5-BU 处理，即能诱导回复突变。但是，并非所有的原初突变回复率都完全相同。在用5-BU 再处理的场合，大多数原初突变的突变回复率极低，回复突变频率高者实属凤毛鳞角。从这些观察或许可以认为：最初用5-BU 诱导的有 GC→AT 和 AT→GC 二种类型的转换，其中一类转换的频率（指 GC-AT）要比另一种类转换（指 AT-GC）的频率显著高出许多。这一结论的证据尚不足令人信服，所以下面不妨进一步来考查一下其他某些专一性已知的诱变剂效应。

另一种比之5-BU 诱变效应专一性程度更高的诱变剂是羟胺，这是一种可与胞嘧啶复合的小分子化合物（图 86）。当羟胺与胞嘧啶结合之后，胞嘧啶即与腺嘌呤配对而不同鸟嘌呤配对。在

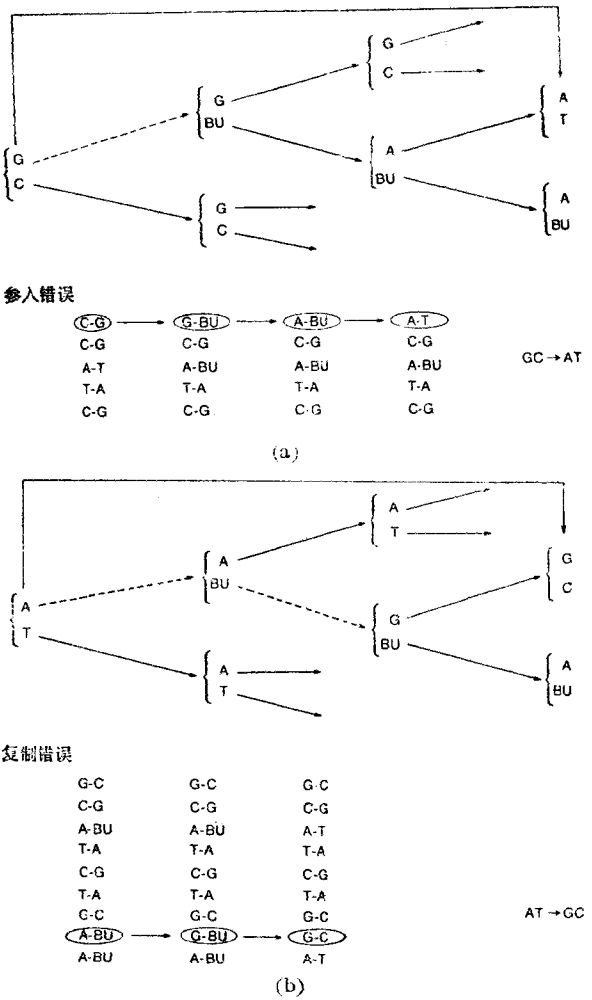


图 85　由参入错误或复制错误产生的突变

(a) 由于参入错误导致GC→AT的转换。一般认为这种转换是由胸腺嘧啶的替代物以烯醇式 5-BU 的形式被结合进原应为胞嘧啶的位置发生的（此时的配对方式是 BU-G)。

(b) 由 5-BU 置换胸腺嘧啶而产生的复制错误与之相反。这里，酮式 5-BU 先与腺嘌呤配对，当酮式 5-BU 链重新复制时，5-BU 分子即从酮式转变为烯醇式，并同时进行碱基配对。由于烯醇式 5-BU只能与鸟嘌呤配对，从而最终导致原AT对转变为 GC对。(引自 E. Freese。)

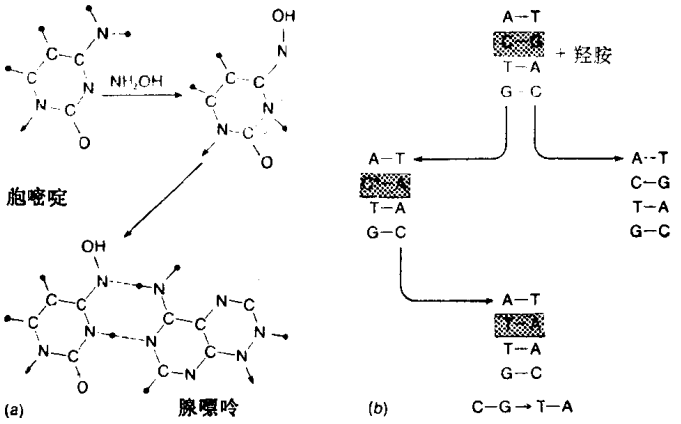


图 86　羟胺诱导突变的机理

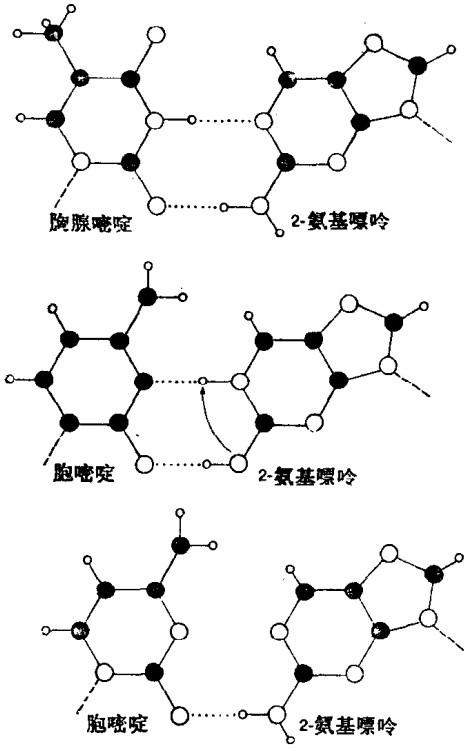
(a) 羟胺对胞嘧啶的作用

(b) 羟胺产生CG→TA转换。其机理解释如下：羟胺同胞嘧啶复合后，形成的羟胺-胞嘧啶复合物专门同腺嘌呤配对，而不与鸟嘌呤配对。

下一轮复制中，读者想象得出，CA 碱基对将会产生一个正常的 CG 碱基对和一个变型的 AT 碱基对；亦即说，羟胺诱导了一类 CG→TA(相当于 GC→AT) 的转换。倘若将由羟胺诱导的突变再置于羟胺诱变剂中用以测定回复突变时，结果无一出现。这一事实支持了下述看法：羟胺诱导的转换只有一种形式，即只能单向地从 CG→TA 转换。

如果用 5-BU 诱导突变之后又继而同羟胺接触，结果证明凡是用 5-BU 再处理时，回复突变率高的突变型(指 AT→GC)亦能在羟胺存在的条件下有效地发生回复突变。这一发现同下述实验结论不谋而合，即大多数在羟胺存在时不发生回复突变的由 5-BU 诱导的突变正好属于 GC→AT 的转换。而在 5-BU 存在时偶而发生回复突变的同样也是属于用羟胺处理不能回复的 GC→AT 转换突变型（因为用 5-BU 诱变时，GC→AT 方向的突变高于 AT→GC 突变——译注)。

这一原理还用来指导其他诱变剂的诱导实验。例如，有一种 2-氨基嘌呤(2-AP)，为腺嘌呤类似物，可在羟胺诱导的以及为 5-



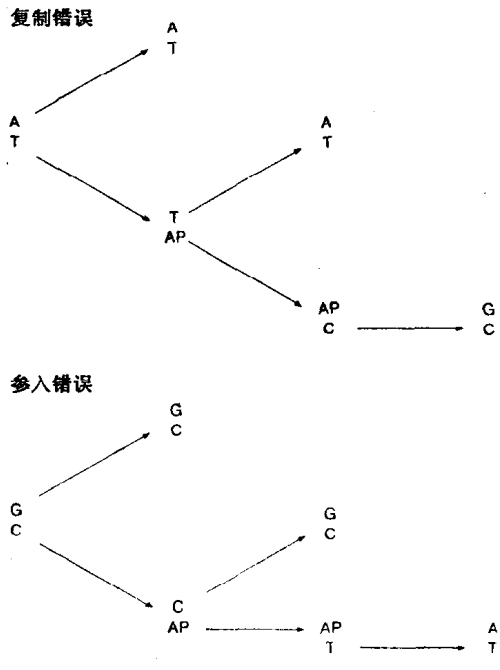


图 87　2-氨基嘌呤诱导突变的机理

图示2-AP诱导突变的二种可能机制。2-AP参入进新链DNA的比例较之5-BU要低，这暗示被参入的2-AP百分比过多时会扰乱DNA多聚酶与DNA模板之间的协调关系而影响到复制进行产生突变。胞嘧啶的复制错误造成AT→GC转换，而参入错误将产生GC→AT转换。

BU 诱导的一类突变型中诱导回复突变。但 2-AP 诱导的大多数突变型却不能在 2-AP 存在时发生回复突变。由此可以推论，2-AP 主要诱导 AT→GC 转换，开始 2-AP 取代腺嘌呤，然后在复制时它再同胞嘧啶配对(图 87)。2-AP 参入 DNA 不太容易，已发现参入只限于整个 DNA 分子长度中一些极为有限的位置。参入困难的情况暗示，一旦 2-AP 参入 DNA 分子，它会改变模板和聚合酶之间结构上的协调关系从而妨碍复制继续。如果确实如此，那么 2-AP 诱导的许多突变很可能均产生于它的结合进 DNA 分子之后引起的复制错误。

亚硝酸是改变 DNA 处于非复制状态下（指非外源分子参入时期）的化学结构的一种有效诱变剂。亚硝酸可以直接脱去含氮

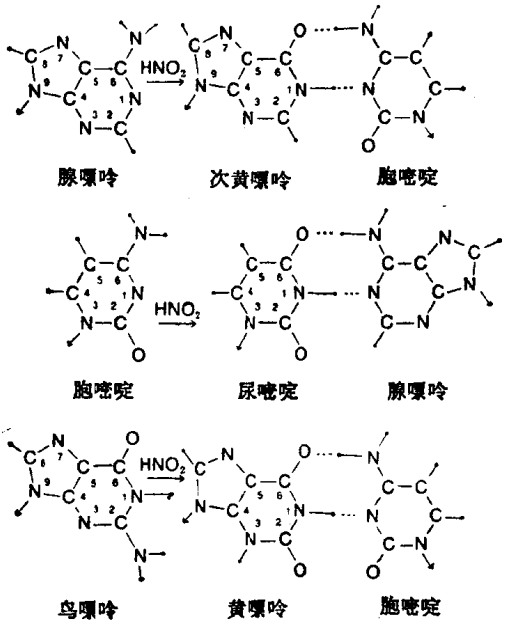


图 88　亚硝酸诱导突变的机理

亚硝酸使腺嘌呤变成次黄嘌呤，次黄嘌呤只同胞嘧啶配对而不同胸腺嘧啶配对(产生AT→GC转换)；亚硝酸又可使胞嘧啶变成尿嘧啶，并使之同腺嘌呤配对，从而取代本应为胞嘧啶与鸟嘌呤的配对(造成CG→AT转换)；至于鸟嘌呤变成黄嘌呤后，仍同鸟嘌呤一样同胞嘧啶配对。由此可以推断，倘若鸟嘌呤转变成黄嘌呤不会产生突变。

碱基上的氨基基团，从而使腺嘌呤转变成次黄嘌呤，使鸟嘌呤转变成黄嘌呤，以及使胞嘧啶转变为尿嘧啶(图 88)。因为次黄嘌呤能同胞嘧啶配对，所以腺嘌呤变成次黄嘌呤后可预计会产生 AT→GC 转换。而尿嘧啶可与腺嘌呤配对，故胞嘧啶变成尿嘧啶后亦可期待会发生 GC→AT 转换。倘若这些推测是正确的话，那么亚硝酸将能逆转由任何碱基类似物或者羟胺诱导的突变。表 8 所列数据支持了这一推测。

表 8　采用化学诱变剂诱导回复突变的分析

突变诱变剂	回复突变诱变剂*				
	5-BU	2 AP	HNO_2	HA	EES
5-BU	±	+	+	−	−
2-AP	+	±	+	+	+
HNO_2	+	+	+	+	+
HA(次黄嘌呤)	−	+	+	−	−
EES(乙基乙烷磺酸盐)	−	+	+	−	±

突变诱变剂	突变	类型
5-BU	GC⟶AT	转换
	AT⟶GC	转换
2-AP	AT⟶GC	转换
	GC⟶AT	转换
HNO_2	AT⟶GC	转换
	GC⟶AT	转换
HA	GC⟶AT	转换
EES	GC⟶AT	转换
	GC⇌CG	颠换
	GC⟶TA	颠换

* ＋＝全部或大多数突变型可被回复；±＝少数突变型可被回复；－＝不能回复。

由于观察到另一类诱变剂所诱导的某些突变在亚硝酸、羟胺或者其他碱基类似物存在时不能发生回复突变，因而引起了广泛注意。这类诱变剂主要为烷化剂。其中有二种为人们所常见，一种为甲基磺酸乙酯 (EMS)，一种为甲基磺酸甲酯 (MMS)。推测此类诱变剂的作用方式远比阐释碱基类似物和亚硝酸的诱变机制来得困难。尽管如此，人们还是知道 DNA 的烷基化主要发生在鸟嘌呤的第七位氮原子上，也可以发生在腺嘌呤的同一位置上，但比较罕见，另外，烷基化作用还能在 DNA 的磷酸骨架主链上发生(图 8))。虽然尚未得知烷化剂如何从 DNA 上脱去嘌呤的细节，

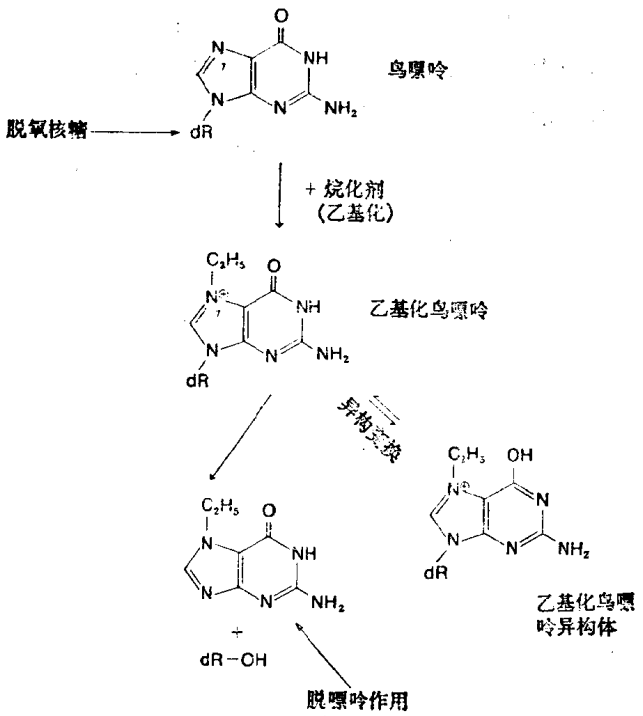
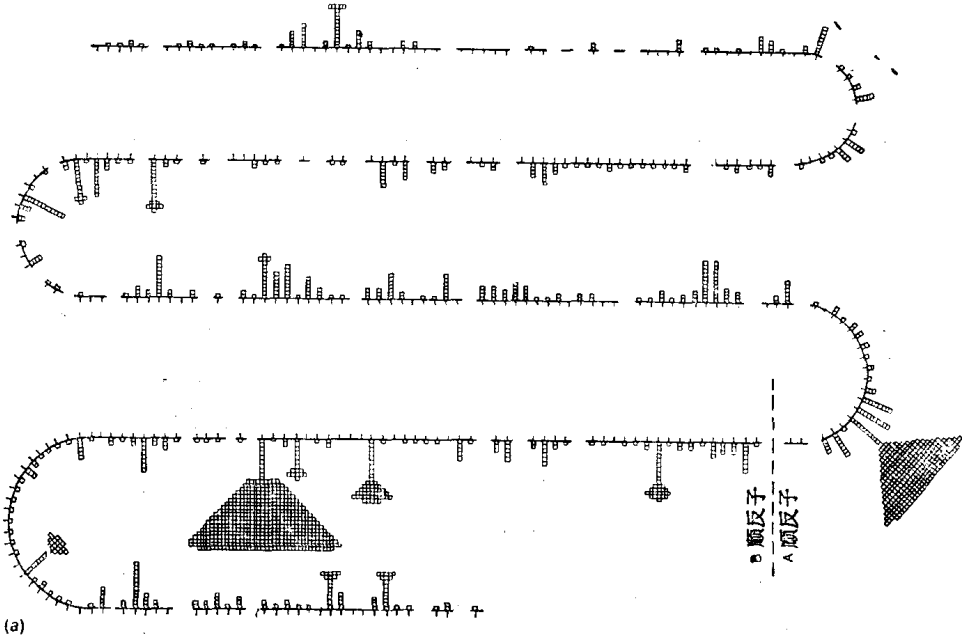


图 89　烷化剂诱导突变机制

图示烷化剂形成突变的作用机制。烷化剂可优先在鸟嘌呤的第七位氮原子 (N-7) 上出现，结果造成脱嘌呤作用。由此而缺失的碱基可由各种嘌呤取代或各种嘧啶替换。若嘧啶取而代之即产生颠换，这类颠换突变一般不可指望以后用诱导转换的诱变剂来使之回复。

但是据信它们必定如此行事。一当嘌呤脱空，四种碱基中的任意一种碱基都能够在下一轮复制中结合进这一空位。因此，烷化剂可诱导碱基对的转换以及诱导嘌呤取代嘧啶或者嘧啶替换嘌呤。后二种置换类型一般称颠换。因此可以推想，一方面烷化剂可脱去鸟嘌呤，而且胸腺嘧啶又可能插入到与此空位相对的位置上，而使其他任何一种诱变剂诱导的 AT→GC 转换均有可能使用烷化剂使之回复突变。这一设想现已得到一些实验资料的支持。

采用上述所有这些诱变剂将有可能设计出一种试验方案，使得在DNA分子的每一对碱基位置上发生一个突变，随后又可用另一种特定的诱变剂，使每一突变回复，从而判断该碱基对的种类。倘若功到事成，则在理论上这种单基因分析法将与碱基顺序分析法的效果不相上下。在研究出一套化学方法用以确定 DNA 排列顺序的技术之前，一般是采用最近邻分析法和诱导突变发生的方法来测定碱基排列的大致顺序的。将诱导突变的方法应用到分析



(a)

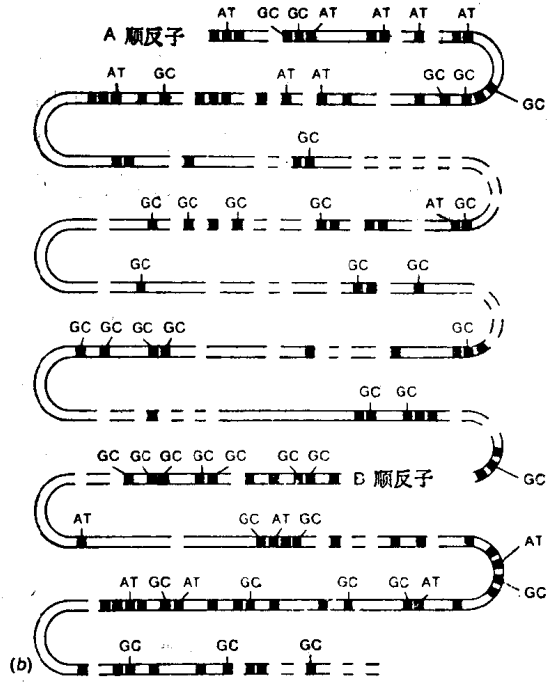


图 90　使用诱变剂确定 DNA 顺序

(a) 图示 T_4 噬菌体 r_{II} 区遗传学图。突变位置在图中用刻线标明。每一小方格代表发生的一次自发突变。没有小方格的位置可通过诱导突变加以鉴定。从图中可见，有二个位点自然突变极易发生。(引自 Benzer, 1962。)

(b) 图示通过突变方式已确定碱基顺序的 r_{II} 区域。许多突变位置由已知其专一性的一种或多种诱变剂诱导出回复突变。根据这些实验推测，现将该区段原先野生型DNA的碱基顺序标出如图。(引自 Benzer, 1962。)

特定基因的排列顺序方面并且成功完成这一独一无二的重大尝试的是 Seymour Benzer。Benzer 曾获得数以百计的 T_4 噬菌体突变品系，它们全部是在 T_4 噬菌体染色体 r_{II} 区的A和B二个基因中发生的突变型(图 90)。虽然在 r_{II} 区中有一小段染色体已经使用这一方法定出了其碱基排列顺序，但由于诱变剂的专一性不甚理想，这一段顺序多少还有点模糊不清。

这里要讨论的最后一类诱变剂为吖啶染料，特别是二氨基吖啶和吖啶黄。其作用方式不同于其他任何一种诱变剂。染料分子

本身可以插在核苷酸碱基配对中由氢键连成的平面环之间，并以非共价键的方式与 DNA 分子复合（图 91）。而且凡由吖啶类诱变剂诱导的突变一般都不能被任何其他的诱变剂所回复，除非是使用另一类吖啶染料。反之，当某种突变可以由吖啶染料诱导回复时，即可断定这种突变在性质上决不会是象前面谈到的那些诱变剂所引起的碱基替换。

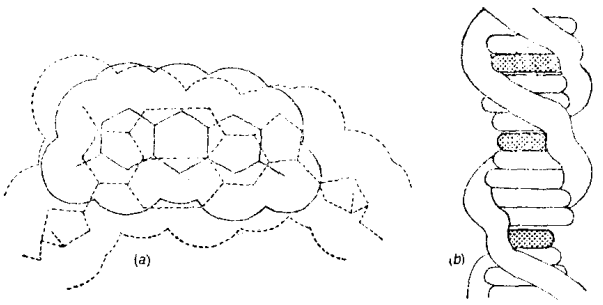


图 91　由吖啶染料引起的突变方式

(a) 二氨基吖啶是发现有诱变作用的吖啶染料之一，已经证明它具有平面型的分子构型，使其能够插在 DNA 分子的碱基对之间。图中所示为DNA双螺旋的顶视图；虚线表示一对碱基位于吖啶分子(以粗线表示)结构的下面。(引自 Lerman, 1964。)

(b) 这是插入 DNA 碱基对之间的吖啶分子侧面观。请注意插入区域中碱基对之间的相对距离。(引自 Drake 1970,及 Lerman, 1964。)

吖啶分子插入双螺旋碱基对之间后，随着复制的进行可导致一对或多对碱基的增加或者缺失（图 92）。假如增加一对碱基(+突变)，并同时伴随有另一对碱基缺失(－突变)的事件发生，则第二次突变就有可能使第一次突变得到校正。碱基对的增加或者缺失都能造成遗传信息的读码移位(即移码突变——译注)。＋突变使读码节段朝前移动，－突变则使读码节段朝后移动。因为＋突变只能由随之发生的－突变来进行校正，反过来，也同样如此。在 T_4 噬菌体中有一个独特的有关 T_4 染色体 r_{II} 区段的缺失突变型，通过诱发与该突变型缺失区段紧密靠近的某些突变即可以说明这一类移码现象(图 93)。T_4 染色体 r_{II} 区由二个基因，即A和B组

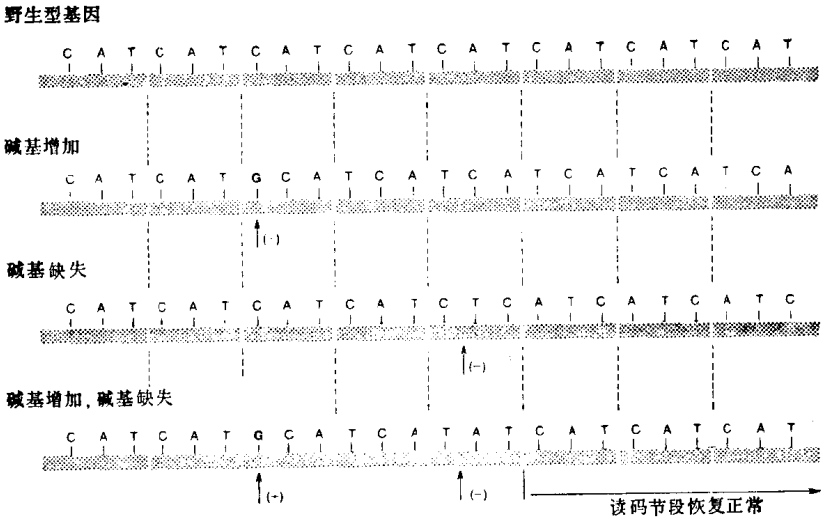


图 92　移码突变

DNA 分子中碱基对的增加或者缺失其所产生的突变效应将会导致遗传信息读码节段的移位。在本图所示的模型中，正确的读码是 CAT。由于增加了一对碱基使读码移位到 TCA，倘若缺失一对碱基，则使读码移位到 ATC。在碱基对增加之后如伴随有另一碱基对缺失(反过来也一样)，便可使读码节段再回复为 CAT。

成。已发现有一缺失涉及A基因的某一区段以及B基因的某一小段。尽管B基因中缺少了一小段，但它的剩余部份仍然具有功能。而A基因由于片断缺失致使功能丧失。

此时这类缺失突变型便可以成为另一种别具一格的可借以检测是否有突变发生的分析手段。假如有某些移码突变发生在无功能的A基因之中，通过观察与之毗连的B基因功能的变化即可把它检查出来；亦即无功能的A基因中移码突变可造成B基因功能的丧失。如果在A基因中发现一个＋突变，则只需要在A基因的不同位点上同时发生一个－突变就可恢复B基因的正常功能。但是，一个＋突变不能校正另一个＋突变产生的效应 (图 93)。(关于阅读 DNA 分子中遗传信息的机理，通过这些突变分析亦可从中获得某些线索，这方面的内容将在第四章遗传信息的转录一节中，详细加以讨论)。实际上，通过遗传学测定还不能明确无误地

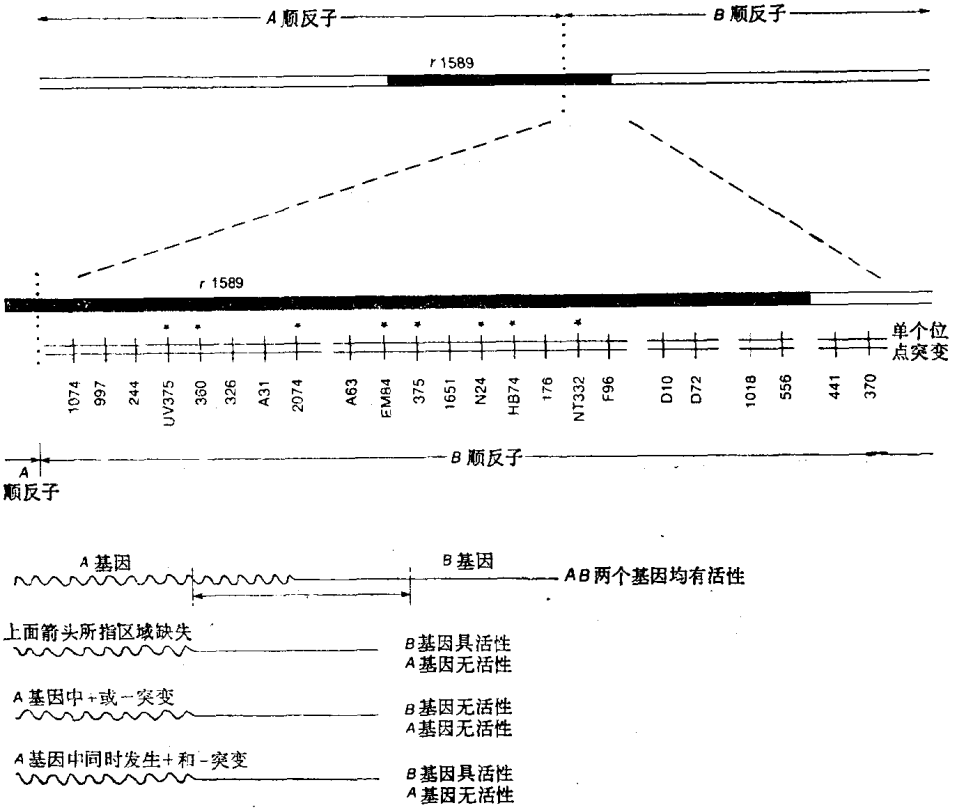


图 93 T_4噬菌体 r_{II}区中在A顺反子和B顺反子相连结的部位发生的缺失突变(黑条即为缺失片断——译注)

图中 r_{II}区的缺失区段涉及到A、B二个顺反子,缺失结果造成A顺反子功能丧失,但B顺反子仍然具有活性。在此种情况下,通过观察B 顺反子的功能表现即可检测出在无功能的A顺反子中碱基对有无增加或是缺失,这是因为这二个相连基因在读码分节时是象一个基因那样从头到尾一气呵成的。由图中可见缺失区域较宽(黑条部分),其中包括多个能发生点突变的位点。尽管B顺反子上有许多碱基对缺失,但它仍有活性;可是在其他该区段没有缺失的品系中,只要其间发生一个点突变即可使B基因失活。

鉴别无功能A基因的突变是属于碱基增加呢,或还是为碱基缺失。人们只是从这二种突变显著的功能差异来区别其性质,而将其中某种突变称为+,另一种称为-则完全是属于人为的规定。

借助于以上概括的这一类实验已经弄清了不少突变机制,但

是对于 X-射线、紫外光、加热等等因素诱导的突变所产生的化学变化本质，人们仍然所知无几。有许多诱变剂都是以非特异性的方式改变着 DNA 的结构，毫无疑问这里必定会涉及到碱基对的替换、缺失以及增加等诸种形式。目前，仅知有少数的诱变剂具有专一性作用，从而可以据以推测在此类诱变剂作用下所发生的分子变化情况。到目前为止，已知有些专一性的诱变剂作用可造成 GC→AT 类型的转换，有些既可造成 GC→AT 的转换又可造成 AT→GC 的转换，有的则转换和颠换二种形式兼而有之。有的诱变剂能够专门引致 DNA 链中个别碱基对缺失或者增加等等，不一而足。借助于所有这些具有特异性和半特异性诱变剂的不同组合的诱导实验，即可获得有关 DNA 碱基排列的大量资料。但由于近来研究核酸顺序的化学分析方法已经取得突破，不免使上述方法显得相形见拙。但不管怎样，这一类工作在揭示突变过程的化学变化方面所积累的丰富资料一直具有很大价值。

重　组

在整个遗传学的发展历史中，尽管重组分析一直占据着这一学科的中心位置，也尽管有关重组机制的本身迫切需要澄清，但令人惊奇的是旨在解释这种现象的理论仍属片鳞只爪，其中，部分原因是在于要设计一个验证重组模型的实验十分困难，但这始终是一个引人入胜的课题。(一种模型是否真有价值，往往取决于预先提出的假设以及依此而设计的具体实验)。

本世纪三十年代，人们提出了二种彼此稍有些不同的旨在解释交换的一般模型，这是迄今为止仅有的二种受人重视的模型(图94)。其中一种称为断裂重接模型，该模型被解释为同源染色体中的一条染色单体和另一条同源染色体中的一条染色单体之间的DNA发生准确的相互交换，先是二条染色单体在同源位点上发生断裂，然后互相交换断头再行重接。另一种模型称为复制选择。其对交换机理的解释如下：新染色单体合成的同时发生重组，此时亲本染色体起着模板作用，新染色单体依次而被合成。倘若亲本染

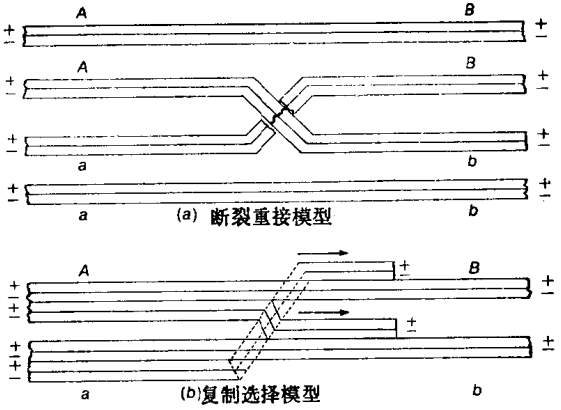


图 94 关于遗传重组的断裂重接和复制选择模型

(a) 断裂重接模型将交换解释为同源染色体的染色单体之间相互进行物质交换。

(b) 复制选择模型则将交换解释为复制错误，即在复制时，一条新形成的染色单体先从一个同源染色单体复制，然后转换到另一同源染色单体复制。另一条新形成的染色单体亦在同一位点上发生转换复制，以这种方式复制的染色单体都含有二条同源染色单体的遗传成分。即是说，每条新合成的染色单体都含有从二条同源染色体复制的区段。

色(单)体靠得很近，则正在合成的新染色单体可以转换模板，先从某一亲本复制，随后从另一亲本复制。这样新合成的染色单体亦可显示出含有同源染色单体之间交换了的物质片断；因此新合成的染色单体将同样携有来自二个亲本染色单体的遗传信息。

从 DNA 分子结构的角度来检测，人们已经设计了好些实验来验证这些模型的可靠性。以 DNA 分子结构理论为依据，可以作出种种推理。Meselson 和Jean Weigle 曾据以设计了一个实验，他们采用的技术方案同早期 Meselson 和 Stahl 用来决定 DNA 复制是否属于半保留型的方法大同小异。

Meselson 和 Weigle 的实验目的是想阐明染色体重组的机制究竟是属于复制选择呢，还是属于断裂重接。试验中以重组 DNA 分子中亲本 DNA (实验中为重链) 与子代 DNA (实验中为轻链)的比例(图 95) 作为判定遗传重组方式的指标。从断裂重接模型推断，重组分子中亲代 DNA(重链)所占的百分比可从 0%

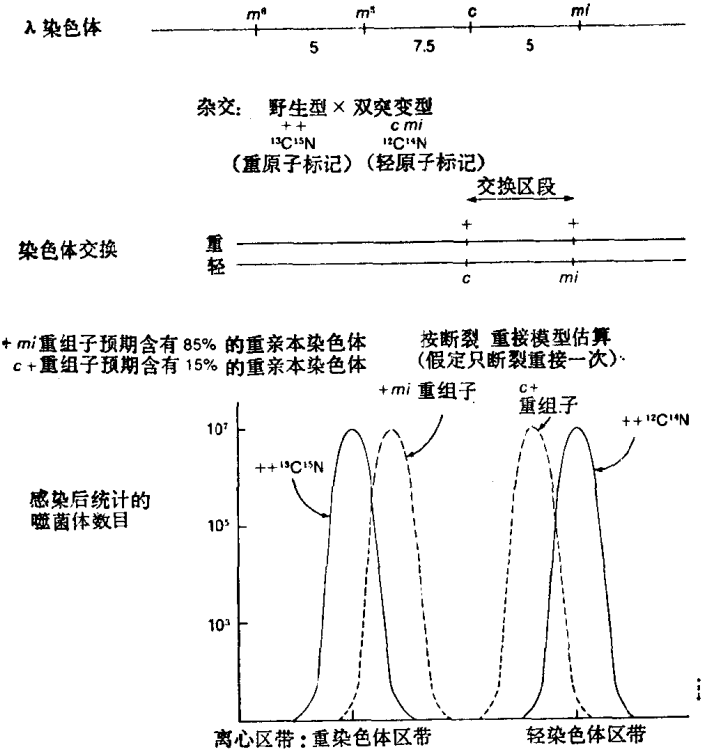


图 95　用于判断复制选择和断裂重接的实验设计

Meselson 和 Weigle 的实验可概括如下：以 λ 噬菌体染色体上已知的基因作为显示有未发生交换的标志基因。因基因 c 和基因 mi 都位于 λ 噬菌体的一端，如果发生重组即为DNA 已经发生交换的明证。在野生型(++)和双突变型(cmi)品系之间杂交，++染色体中全部用引进重原子$^{13}C^{15}N$ 标记。$c\,mi$ 的染色体则为一般的轻原子 $^{12}C^{14}N$。假若交换系起因于断裂重接，那么通过 c和 mi 之间的交换可预期得到二组重组子，其中一组为+ mi 重组子，其染色体上有 85%的区段为重原子的 DNA，与之相应 c+重组子上只有 15%的区段为重原子DNA，反之倘若交换的机理为复制选择模型，则重组子染色体决不会超过 50%，这是因为双螺旋分子中有一条链是在轻原子培养基条件下重新合成的。在实验中将感染的大肠杆菌离心沉淀获得不同 DNA组份；再进一步用遗传学方法测定各个不同DNA组份的基因成分 (以便确认发生交换的重组子与未交换的非重组子——译注)。

至几乎 100%，依 DNA 断裂重组的次数而异。而在复制选择模型中，由于重组是在新 DNA 复制过程中而产生的，因此，重组子中亲代 DNA(重链)部分决不可能超过 50%。可是在重组子染色

体中发生亲本 DNA 片段超过 50% 的情况下，只能告诉人们断裂重接确已发生，但并不能排除同样也有复制选择出现的可能性，因而使该项实验分辨模型的能力受到了限制。

Weselson 和 Weigle 在这项实验中以较易分离的 λ 噬菌体 DNA 为材料，其分子量是如此之小，以致于通过离心就很容易将重型(用同位素标记)，中间型(有部分标记)及轻型(未标记)三种 DNA 分子分开。除此之外，早期用 λ 噬菌体为材料所做的工作亦提供了一个很派用场的该噬菌体染色体遗传学图（图 95)。由图可见，标志基因 *c* 和 *mi* 系位于在λ噬菌体染色体 DNA 的右端，这就有可能通过重组产生出携带有亲本 DNA 区域（实验中为标记重链——译注）与子代 DNA 区域（实验中为未标记轻链——译注)的比率各不相同的重组子染色体(这是因为发生交换的区域不同 ——译注)。

如果二个亲代染色体分别为 $c\ mi^{12}C^{14}N$ (较轻的 DNA)和 $+\ +^{13}C^{15}N$(较重的 DNA)，那么在二个基因位点之间发生断裂重接而产生的重组，则 $+mi$ 重组子携带的较重 DNA 部分必然要比 $c+$ 携带的要多(因 *c mi* 位于一端——译注)，图 96 所示即为该杂交实验的结果。这项实验虽然做起来比较容易，但得到的结果不太明确。产生这一困难的原因在于：噬菌体的增殖不是同步的，并且除了使用重复感染技术之外，很难掌握操纵杂交的进程。这是因为如果每一种亲本噬菌体都只有一条染色体进入一个寄主细胞，那么除非经过一轮或几轮复制后产生大量的个体，否则重组不太可能发生；然而矛盾在于只要发生一轮复制，重原子标记过的亲本染色体就将化为二个密度介于重轻链之间的中间型子代 DNA 分子而致面目全非。基于这一理由，实验必需要如此处理：即采用大量重复感染，增加二种亲本噬菌体的个体数，以便提高亲本染色体之间重组的机会。否则，由于发生重组之前，重组 DNA 亲本已预先进行了增殖，结果使得重型 DNA 的重组子(指在重组子中重型亲本 DNA 占 50% 以上的重组子) 含量大为冲淡，密度分析曲线中的重型 DNA 的峰亦相应平缓，结果必然会造成实验的分辨

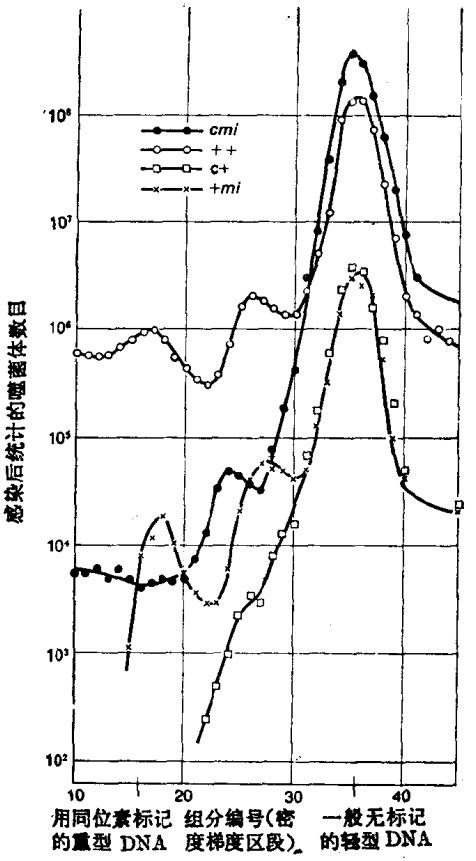


图 96 Meselson-Weigle 的实验结果

图中结果是支持关于解释断裂重接模型假说的。如图 95 所推测的那样，在某些 $+mi$ 重组子中其DNA分子有 85% 系为标记的重型DNA，而在其他各类 $+mi$ 重组子(中间型的 $+mi$ DNA以及轻型的 $+mi$ DNA)中，则无法据以判断选择复制和断裂重接二种模型孰是孰非。(引自Meselson 和 Weigle, 1961。)

力下降。因此在根据所有重组子的分子密度来考虑所得数据时，必须记住这一点。

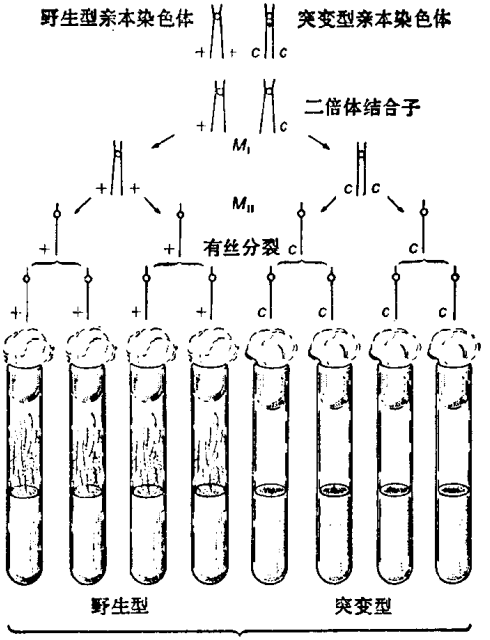
如果是基于断裂产生 $+mi$ 重组子的话，由于这类重组子据推测应含有 50% 以上重 DNA，所以它在实验分析中具有特殊意义。经测定证明 $+mi$ DNA 出现三种密度高峰。一种峰几乎全为

重 DNA，另一种峰密度介于轻重之间，第三种峰为轻 DNA。标记的重 DNA 分子的存在为断裂重接模型提供了有力证据。中间型 DNA 分子及轻型 DNA 分子则被解释为是这一过程的产物，即在发生交换(重组)之前，染色体已经历了一轮、二轮或更多轮的复制，因而标记重链部分正渐渐趋向减少。而由相互交换(重组)得到的另一类 $c+$ 重组子，据推测也应携带一小段标记重型 DNA 分子(约占整个DNA 分子的 15%)，如图 96 所示，在轻型 DNA 峰靠向重型 DNA 峰一侧，在密度曲线中间确实出现了一个与预期相应的肩形波峰。

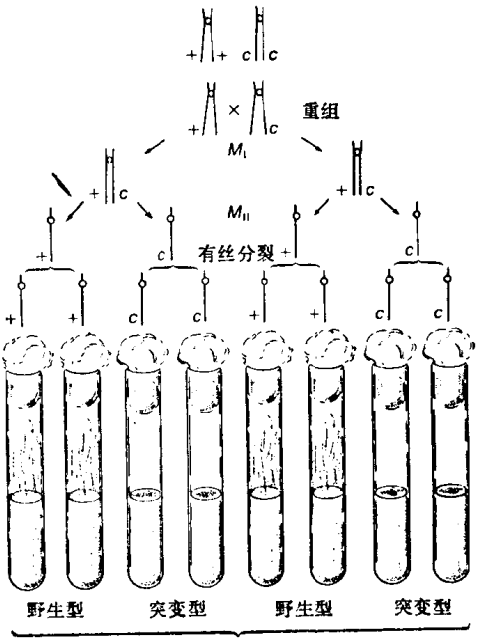
关于上述试验的另一个不确定因素是来自于这一疑问，即由于在重型 DNA 分子中可能有少量新合成的轻型 DNA 分子存在，是否会使重型$+mi$ DNA 重组子要比所预测的含量(85%)稍许少些。为了寻求这一问题的答案，进行了一些重复实验；不过在第二次实验中用于杂交的二种亲本噬菌体都携带有同位素重原子标记的 DNA。倘若重组子的产生确实是出于断裂重接方式，根据半保留复制将会使重组子中重 DNA 成分含量减半而冲淡的估计(指 DNA 先复制完毕以后再发生断裂重接——译注)，可以推想由之必然会获得重型、轻型、中间型三种类型$+mi$ 重组子。实验资料表明，在某些重组子中，虽然其密度结果同重型亲本 DNA 分子极为相近，但二者亦有不尽符合之处，这使人们猜测是否在与断裂重接进程的同时，也会出现有某种修补合成。从这些资料来看，既然重组亦伴随着复制，便不能说一定不发生复制选择。目前，一般认为，在 λ 噬菌体染色体的实验中，尽管尚有某些似是而非之处，但通过断裂重接伴随有遗传重组的进程，这是无可置疑的。

已统计的大多数遗传重组分析都未涉及到揭示重组的分子机制和化学变化，只是按照传统的经典遗传学的方法行事。即使如此，人们从这些分析仍然了解到不少有关遗传重组的重要事实。此类方法尤其适合在某些真菌中作重组分析。

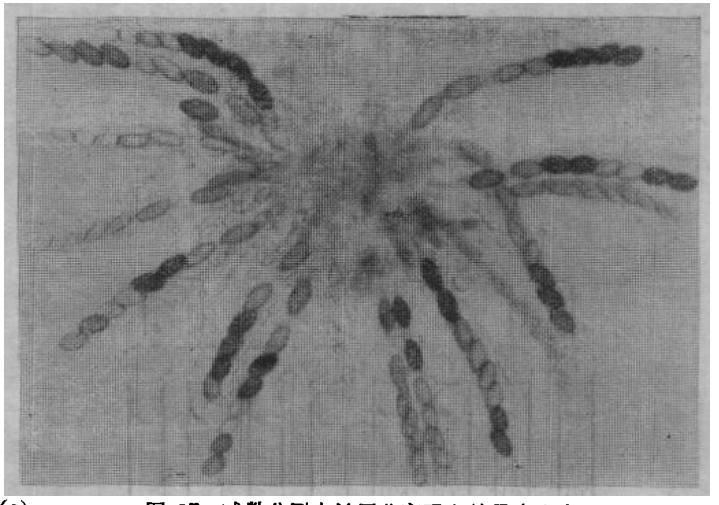
读者可以回顾一下前已述及的孟德尔的等位基因分离法则。它的形成并非出于对分离机理的解释而是来自于对分离结果所作


野生型亲本染色体
突变型亲本染色体
二倍体结合子
M_I
M_{II}
有丝分裂
野生型
突变型
4:4 分离子囊类型

(a)


重组
M_I
M_{II}
有丝分裂
野生型
突变型
野生型
突变型
2:2:2:2 分离子囊类型

(b)



(c)

图 97　减数分裂中关于分离现象的目击观察

减数分裂一般连续分裂二次。图示第一次分裂(M_I)为同源染色体相互斥离。第二次分裂(M_{II})为染色单体相互斥离。图中为野生型(+)亲本和突变系亲本(*c*)的一对基因杂交。

(a) 图示第一次减数分裂(M_I)基因＋和基因 *c* 分离，在这里同源染色单体在基因 *c* 到着丝粒之间的区域内并未发生交换，因之基因*c*和基因＋在M_I分离时，产生的子囊型为 4(野生型):4(突变型)。(引自 Beadle, 1947。)

(b) 图中基因 *c* 和基因＋的分离发生在 M_{II}，因为同源染色单体在着丝粒和基因*c*之间的区域发生一次交换，因此在第二次减数分裂(M_{II})之前，每条染色体便都共同带有基因＋和基因 *c*。通过第二次分离得到 2:2:2:2 的子囊孢子类型。

(c) 这是从一个子囊里得到的许多子囊孢子的光学显微照相。在六种可能的分离类型中图中可见到五种，黑色孢子为正常型(+)，浅色孢子为白色突变型 (*wc*)。(引自 R. L. Phillips 与 A. M. Srb, canJ Genet Cytol, 1967。)

的统计学概念的归纳。通常可以从一个测交试验的结果分析，就可以简要地说明有关分离法则的基本内容，以及为论证这一法则所习惯采用的统计学分析方法。测交试验在设计上是统计测交亲本所产生的二种类型的配子数以估算这两种配子所产生的频率，并在这些资料的基础上进一步阐明分离现象。在二倍体生物中，测交组合 $Aa \times aa$ 产生二种子代，即 *Aa* 和 *aa*，二者比例接近 1:1。这是因为 *Aa* 亲本产生二种配子，一种配子携带有等位基因 *A* 和另一种配子携带有等位基因 *a*。二种配子产生的比例

亦为 $1A : 1a$。人们将这一现象解释为：在配子形成阶段，当细胞减数分裂时 A 同 a 各自分离，最终形成二个 A 配子和二个 a 配子。

Garl Lindegren 用粗糙链孢霉 (*Neurospora crassa*) 为实验材料首先证明：分离乃是伴随每一次减数分裂普遍发生的现象。但是发现也有许多类减数分裂进程其所决定的子囊配子的频率与统计学上的预估并不尽相符合。作者的该项工作有一突出优点，即在链孢霉中，子囊孢子在子囊中的位置排列能够按原先减数分裂过程中它们产生的先后顺序“维持不变”，从而为揭示孢子产生的减数分裂方式的某些详细情节提供了线索(图 97)。

第一次减数分裂形成四个子细胞。在雄性哺乳类动物个体中这四个子细胞全部发育成为精子细胞。在链孢霉中四个子细胞则发育成子囊孢子(随后再行一次有丝分裂产生四对子囊孢子)。雄性哺乳类动物个体中，数以万计的减数分裂过程产生的精子细胞共同聚集在输精管中，人们无法将来源于同一个原始精母细胞的那些精细胞从中分离出来。而在链孢霉中，四个子囊孢子在同一个称为子囊的囊状组织结构中发育，因而每一个子囊实际上都保存了一次减数分裂过程的真实“记录”。

读者从图 97 可以看到，由链孢霉的二个品系之间单因子杂交会产生二种类型的子囊。第一种类型为 *AAaa* 和 *aaAA*，追溯其 A 与 a 的分离时间应发生在减数分裂的第一阶段 (M_I)。另二种类型为 *AaAa* 或 *aAaA*，以及 *AaaA* 或 *aAAa*，其 A 与 a 的分离应发生在减数分裂的第二阶段 (M_{II})。这里有二个常用的概念需要稍加说明：(1) 发生第二次分裂分离的子囊频率是依 A 基因和着丝粒之间距离而变化的（读者从图 97 可以判断，只有当某一减数分裂的全部产物能够无一遗漏地加以鉴别时，才可绘制染色体着丝粒区域遗传图)。(2) 唯有通过交换使等位基因相互换位才会在 M_{II} 发生等位基因的分离。

当 Lewis 在果蝇中发现基因内交换后不久，同一现象亦在链孢霉中接踵出现；并由之还获得另一意外结果——其表现同分离定律大相径庭。例如，倘若在同一基因的二种不同突变型之间做一次杂交，为了叙述方便起见可将这二种不同突变视为非等位基因并以符号来表示该对杂交：$a_1+\times+a_2$，得到的双倍体接合子为 $a_1+/+a_2$。按照经典交换原则将产生下列类型的子囊：a_1+，a_1a_2，$++$，$+a_2$。本例中，a_1 与 $+$（为 a_1 的野生型"等位基因"）分离，a_2 与 $+$ 分离，因而可以预期这两种分离比应均为 1:1。但在实际上，所预测的分离方式从未发现(图 98)。在不少例子中，从子囊类型的观察证明在 $+$ 和 a_2 之间则呈 3:1 分离。由此可见，常规

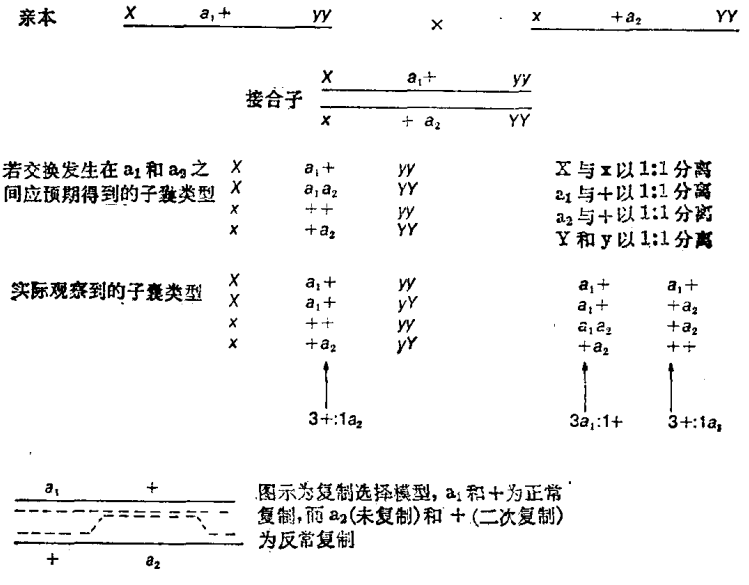


图 98　基因内突变的分离方式

本图表示一对杂交，a_1 和 a_2 为同一基因(a)中不同的突变。a_1 和 a_2 之间发生一次交换预期产生一种属于 a_1+，a_1a_2，$++$，$+a_2$。(4a:4+)组合类型的子囊。由于 a_1 和 a_2 之间的距离较短，故重组型子囊孢子的数目亦必较少，这是一点。可是另一方面，在所观察到的许多重组子子囊孢子中，不约而同地发现有一对基因按 1:1 分离，而另一对基因按 3:1 分离，如图所示。下图所示为用复制选择模型对这一不同寻常的分离方式如何通过重组的解释。

所说的等位基因分离方式并非照例适合于基因内重组。据图 98 所示模型，用复制选择模型却能够解释此类孢子的反常行为。

但是，试图用复制选择模型来解释上述重组过程，还会碰到许多问题。第一，假若新合成的染色单体是按照开始从一个同源染色单体复制，随后又转换到另一个同源染色单体复制的话，那么人们将会注意到，作为模板的染色单体势必为双螺旋 DNA 分子（亦即所谓的染色单体实际上为亲本染色体——译注）。这就意味着，新合成的双螺旋 DNA 是以旧双螺旋 DNA 为模板复制的。显然，这违反了半保留复制法则。因为迄今为止，尚缺乏真核生物中染色体"全保留"复制的实验证据。

第二个问题是时间有矛盾，因为在细胞周史中染色体发生交换的时期同 DNA 合成的时期前后相错，不相重合。在任何一种真核生物细胞的生命周史中，大量的 DNA 合成总是处于 S 期（参见图 16），而同源染色体开始配对则发生在 G_2 后期或 M早期。当 G_2 期将近结束时，核仁中的染色质物质逐渐消失，细长丝状的染色体开始出现（如光学显微镜所见）。同源染色体配对（联会）即在这一时期进行，配对后不久，染色体即显示出已经发生交换的特征（图 99）。

尽管由于存在着全保留复制以及 DNA复制与染色体配对，在时间上的矛盾而使人们难以接受复制选择模型，但要完全抛弃这一模型却也感到为难，因为它在解释链孢霉以及有关真菌的反常子囊类型方面仍然具有一种魔力。保留这一模型的一个主要推动力是来自下面的证据，即在染色体配对时会出现有修补合成，人们揣测这种合成便可能伴同有断裂重接；另外一个原因是，目前已有可能建立由修补合成和并不需要以全保留复制作为前提的复制选择结合起来考虑的模型（图 100）。

基因内重组还揭示出另一条线索，表明过去根据基因间重组而提出的这些过于简单的交换模型很值得推敲。读者可回顾一下，前面有关交换过程的讨论中曾经谈到，在发生多重交换的场合，第一次交换可以发生在整个染色体长度范围内任何位置上，第一次

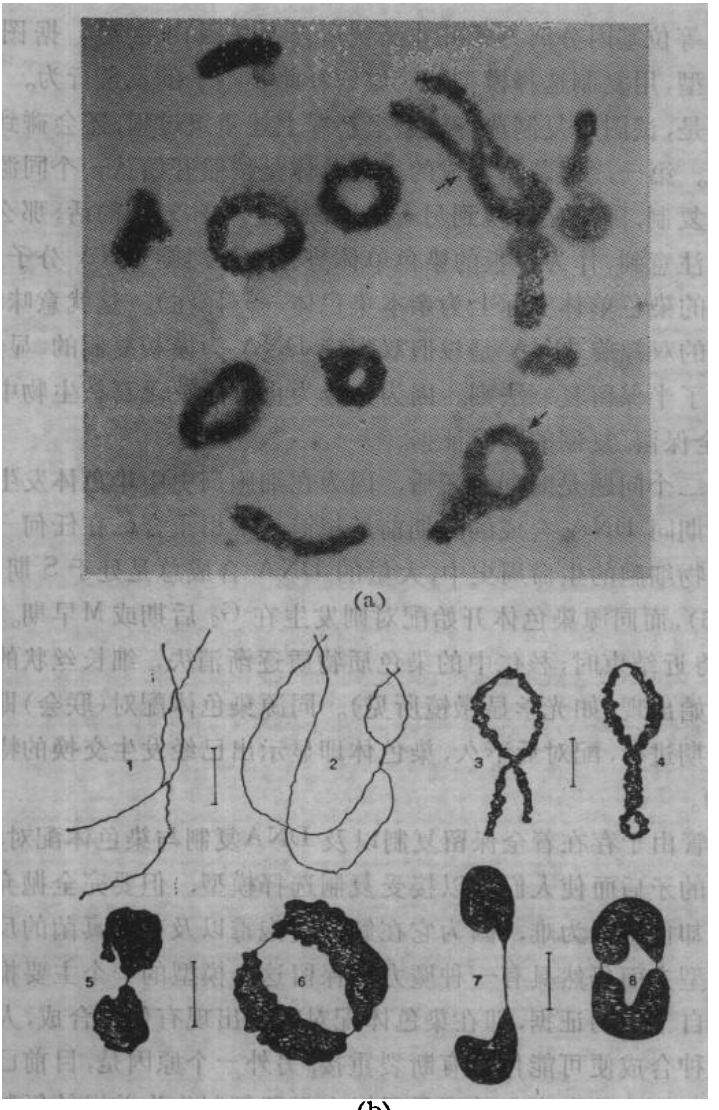


图 99　染色体交叉的光学显微图象

(a) 减数分裂 I(M_I)前期所观察到的在同源染色体之间形成的交叉可以认为是表明染色体发生交换的实物证据，图中为蝗虫雄性个体精母细胞的显微照相，可以肉眼观察到交叉点。(由 R. L. Phillips 与 R. V. Knowles 赠送。)

(b) 这是染色体交叉端化的简略图解，端化起始于早中期，结束于后期(出现于M_I之前和M_I期间)。(引自 A. Henderson。)

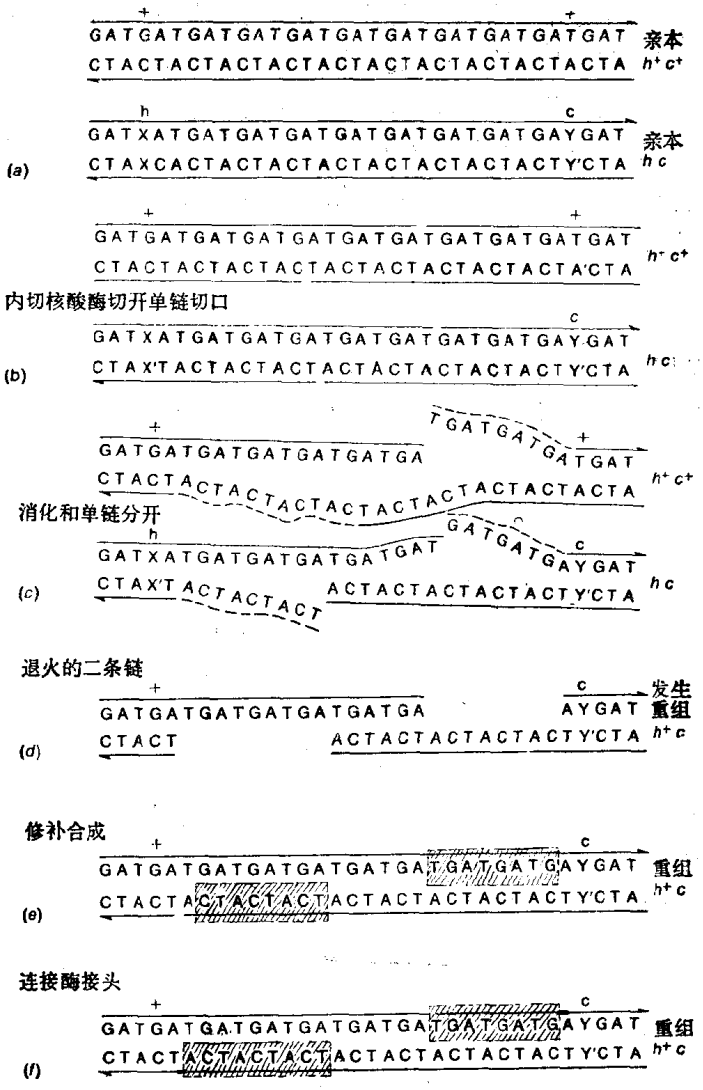


图 100 伴随有限区段 DNA 合成的断裂重接型重组模型

关于重组中涉及断裂重接和有限区段DNA合成的过程可以通过下列各步给以解释：(a)亲本染色体，一条为＋＋，一条为*hc*。 (b) 内切核酸酶将单链切断，图示四条链中各有一个切口。(c) 切口显露后，通过双链解旋和核酸酶消化产生单链区。 (d) 剩余链退火以后的互补区，即二条单链在互补区中又重新通过氢键配对，回复为双链的同时发生重组。 (e) 恢复整个DNA分子正常碱基配对的DNA修补合成。 (f) 由连接酶完成断头连接。

交换一旦发生之后，第二次交换便不太可能紧挨着第一次交换的位置出现。但在距离第一次交换较远的某些位置，却有可能再发生一个第二次交换。为此，有必要对上述概念用数学语言略加说明一二。如果基因 A、B 和 C 相互连锁(图 101)，借助于下述测交试验将能够提供测量这三个基因相互之间距离长短(A-B, B-C 和 A-C)的方法：

$$\frac{ABC}{abc} \times \frac{abc}{abc}$$

证明基因 A 和 B 之间确已发生交换的证据是重组子子代 *Aabbcc*, *aaBbCc*(单交换)和 *aaBbcc*, *AabbCc* (双交换)的出现。在这二种

染色体 A15 B15 C

亲本基因型

测　交　*AaBbCc* × *aabbcc*

子代基因型	每种基因型子代数	A-B区中交换		B-C区中交换		A-C区中交换	
AaBbCc	350	*Aabbcc*	73	*AaBbcc*	73	*AaBbcc*	73
AaBbcc	73	*aaBbCc*	73	*aabbCc*	73	*aabbCc*	73
aabbCc	73	*AabbCc*	4	*AabbCc*	4	*Aabbcc*	73
Aabbcc	73	*aaBbcc*	4	*aaBbcc*	4	*aaBbCc*	73
aaBbCc	73		154		154		292
AabbCc	4						
aaBbcc	4						
aabbcc	350	图距(按重组百分率计)	15.4		15.4		29.2
总　数	1,000	双交换预期值 = 0.154 × 0.154 = 2.37%					
		双交换观察值 = 8/1000 = 0.8%					

图 101　正干涉概念

所谓正干涉是指：第二次交换事件的出现受到第一次交换事件的干涉。按数学语言说，即两个独立事件同时发生的频率为它们各自独立机率之乘积。如果双交换事件频率低于预期值，如本例中(预期值 = 2.25%，观察值 = 0.4%)那样，则表明第一次事件降低了第二次事件出现的机率（正干涉），倘若双交换事件高于预期值，则可以说是第一次事件增加了第二次事件出现的机率(负干涉)。

单交换中，设若重组子子代都可达到总数的15%，这即表示A和B相距为15个图距单位，而B和C亦相距15个图距单位。A和C的距离为AB和BC间距的加和即30个图距单位。

还有另一种测量AC之间距离的方法，即统计在基因A和基因C之间已发生重组的重组子子代比例数，在图101中，这一数值为29.2%。二种方法测出的数值之所以不同，可由下述原因给予解释：在基因A和基因C之间发生了二次交换，产生的"重组子"表型上相似于亲本类型。这一数目(16)并未加进A-C重组子的观察数(292)。本例中，倘若按A-B区和B-C区独立发生交换计算，则实际观察到的双交换重组子数目比预期值要少得多。如果某一区域中出现的一次交换对其近邻区域中另一次交换的机率毫无影响，则这二个区域中同时发生一次交换事件频率将为它们各自机率的乘积，即为0.15×0.15或2.25%。但实际观察到的频率要远比理论估算值为低，在1000个子代群体中，双交换重组仅仅只有四次，亦即0.4%(总的双交换重组子为$2\times0.4\%=0.8\%$)。双交换的实际值比根据机率预测的理论值减少的现象一般即称正干涉。一般来说在测量染色体基因之间相互间距的场合时，发现第一次交换总是毫无例外地会降低邻近区域第二次交换的机率。

但基因内重组(指在双交换中，有一个交换属于基因内交换的场合——译注)并不遵守正干涉法则。事实上甚至可见到与之相反的效应，即某一交换事件的出现会提高另一邻近交换事件的机率(负干涉)。请读者考虑一下Patricia St. Lawrence收集的数据(表9)。她用链孢霉作材料设计了一对杂交，可以同时测定染色体上六个间隔的距离，详情如下：杂交亲本的基因型一为Aq_1+c，与之相对应的另一亲本基因型则为$a+q_2C$，染色体上的六个区段间隔为A-q_1，A-q_2，A-C，q_1-q_2，q_1-C和q_2-C(图102)。除q_1-q_2属于基因内间隔区之外，其余所有的间隔区均属基因之间的间隔区(也包括基因内区段加上基因间的区段，如a-q_2，q_1-c——译注)；也就是说，q_1和q_2是同一基因内相互独立的突变。

在检测上述杂交组合子代类型之前，事先已经标清A和q_1，q_1

表 9　负干涉现象

亲本类型：$Aq_1+c \times a+q_2C$

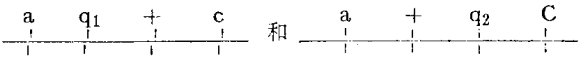
染色体： 

图　距：　5　0.01　5　　5　0.01　5

设选择子代为非烟酸依赖型，即＋＋型，则：

1. 从一次交换产生的子代将为 $a++c$；例如

$$\frac{Aq_1+c}{a+q_2C} \times \quad a++c$$

2. 由双交换产生的子代将为

$A++c$　或　$a++C$

预期它们发生的频率较低（5%×0.01%）。

3. 由三次并发交换产生的子代为 $A++C$ 按正干涉法则其出现频率更低（5%×0.01%×5%）。

4. 然而，下列数据表明，所有＋＋型子代频率大体一致，这意味着在 q 基因内部所发生的双交换和三交换与单交换同样频繁，此即为负干涉。

杂　交　对	表现为＋＋型子代数		
	每10^4子代中出现的个体数	单　交　换	双　交　换
$Aq_1+c\times a+q_2C$	4.3～6.9	39	33
$A+q_2c\times aq_1+C$	4.1～7.5	21	24
用其他 q 突变型做的杂交			
1	4.4～6.5	36	16
2	4.9～6.3	14	7
3	0.8～1.3	17	25

和 q_2，q_2 和 C 之间的图距分别为 5，0.01 和 5。从上述图距即可推算出双交换和三交换的频率，其数值之小简直微乎其微（表 9）。事实上，由于双交换和三交换的预期频率如此之低，以致单单采用常

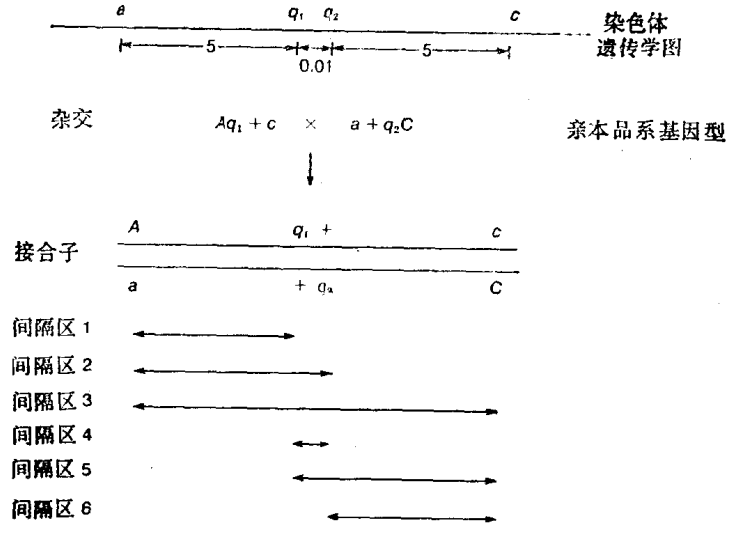


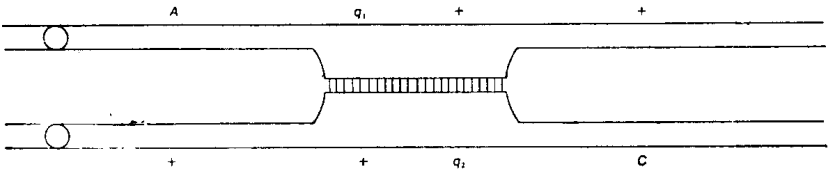
图 102　涉及到基因间和基因内间隔区的一对杂交

图中染色体之六段间距是按杂交后代重组子的频率"求算"的，在这六段间距中，既包含基因间的间距，亦包含基因内的间距，如 q_1-q_2 为基因内间距，a-q_1 为基因间的间距（亦包括基因间与基因内间距之和——译注），a-q_2 则为基因间与基因内间距之和等等。关于负干涉的概念可从下述事实得到支持，即单交换、双交换、三交换的频率基本相同；这与根据各个单交换频率预测的机率全然不同。

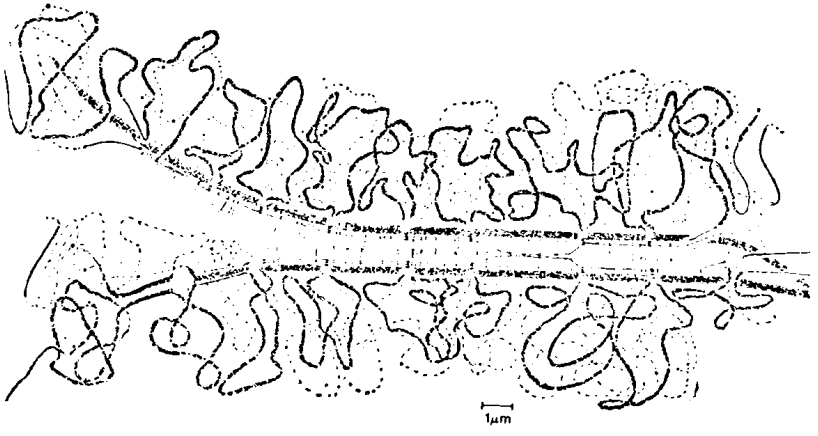
规的计算重组子子代数的方法已无法对之进行有效的检测。好在从此处描述的实验来看，可以通过淘汰属于 q 位点野生型（++子代）以外的所有子代的办法来大大提高实验的分辨能力。因为 q_1+，$+q_2$ 和 q_1q_2 三类子代在表型上完全相同〔本例中，它们均需维生素（烟酸）维持生长〕，借助于对子代类型营养特性的简单检测就可以把基因 q 位上的野生型++区别出来（因野生型在表型上不需要补加烟酸亦能正常生长）。在操作上可将杂交子代的子囊孢子接种于基本培养基上，此时，绝大多数的孢子（主要为非野生型孢子）将被淘汰，在实验中只需要计算在 q 基因上为野生型的双交换与三交换的重组孢子的频率（根据基因 a, q, c 三者关系确定），方法甚为简便，其所得数据列于表 9。

由于长期以来，人们对正干涉现象已经形成了固定的概念，因

此 St. Lawrence 的实验使许多遗传学家感到震惊。负干涉现象告诉人们，与早先的经验恰恰相反，第一次交换事件的存在可以成倍地增加与其近邻的第二次交换事件的发生。后来在许多生物种属中作了基因内重组的研究，结果都观察到有负干涉现象，这一发现也就不再时新好奇了。现在的问题是，正干涉和负干涉实际上



(a) 有效对合区假说(图中与A等位的+，即为 a；与C等位的+，即为 c)



(b)

图 103　说明正干涉和负干涉的一种模型

(a) 高效对合交换区假说：同源染色单体之间在高效对合区段的许多位点上紧密对合，配合得以实现的机率遵循正干涉法则。但一旦紧密对合成功，便在很短的紧密对合区发生许多染色单体之间的交换。在分析包括这样一个区域里发生的双交换重组结果时，看起来便表现为负干涉。

(b) 一个紧密相接的高效对合区域。从图中二条轴芯(由二条染色单体组成)处，可以看到，来自二条染色单体的 DNA 都从浓缩的染色单体核心部位向外延展伸长，形成一种结构上的复合体，而使同源DNA链靠得极近，结果使得由断裂重接方式引起的交换和由 DNA 修补合成及复制选择双重机制引起的重组都能够据以发生。(引自 Moens, 1968。)

是暗示着二种不同性质的交换事件呢，还是这二种干涉根本就是一码事，只不过基因内的交换比起基因间的交换来说，前者揭示了更多的有关交换过程的情况罢了。

关于基因间重组和基因内重组之间所观察到的差别，已提出好些模型给予解释。迄今为止，尚无一种由实验证实的机制能言之有据地解释所有观察到的结果，只有一种模型可将大多数有价值的资料统一归纳起来，这就是所谓的高效对合交换区假说（图103）。该模型的杰出之处是将那些表现负干涉法则的交换区段能够从总体为遵循正干涉法则的交换区段中区分出来。例如，假说提出染色体总体为遵循正干涉的交换法则，但在同源染色体上有一段长度从几百个碱基到几千个碱基对之间的区域其配对方式别具一格，即在一个极短的区段由二条染色单体对合极其紧密，并能够发生多重交换。在此种场合，单体之间所发生的物质交换便可以用断裂重接模型解释，但是在高效对合交换区所发生的交换事件，也同样可以用复制选择有限段的 DNA 合成给予说明。

在前面关于 DNA 一级结构的讨论中，曾经述及 Benzer 的实验证据，作者认为核苷酸碱基对乃是交换的最基本单位。Benzer 以 T_4 噬菌体为材料，力求在染色体 r_{II} 区域诱导尽可能多的突变，使之达到“饱和”；然后将所有突变型进行各种组合杂交测验，以期阐明基本的(交换)重组单位之间可能发生重组的最低频率。读者可以从表 10 找到某些重要的在鉴别 r_{II} 突变型同野生型噬菌体方面行之有效的主体性状，图 104 所示即为 r_{II} 突变型的不同形态特征。此类突变型同样还可以用来分析稀有的重组子，这是因为 r_{II} 突变型在可供野生型噬菌体正常生长的寄主细胞内生长极差或者不易存活，但是，对于另一类不同的寄主品系某些突变型却能安然无恙。换句话说，在技术上可以借助于某一类寄主细胞来检测野生型重组子，而使用另一类寄主细胞又能够发现r_{II}突变型噬菌体品系之间差异(表 10)。

r_{II} 突变型中的互补，遵循链孢霉突变型中前已述及的法则，即A基因中的任一突变型当与B基因中任一突变型感染同一寄主

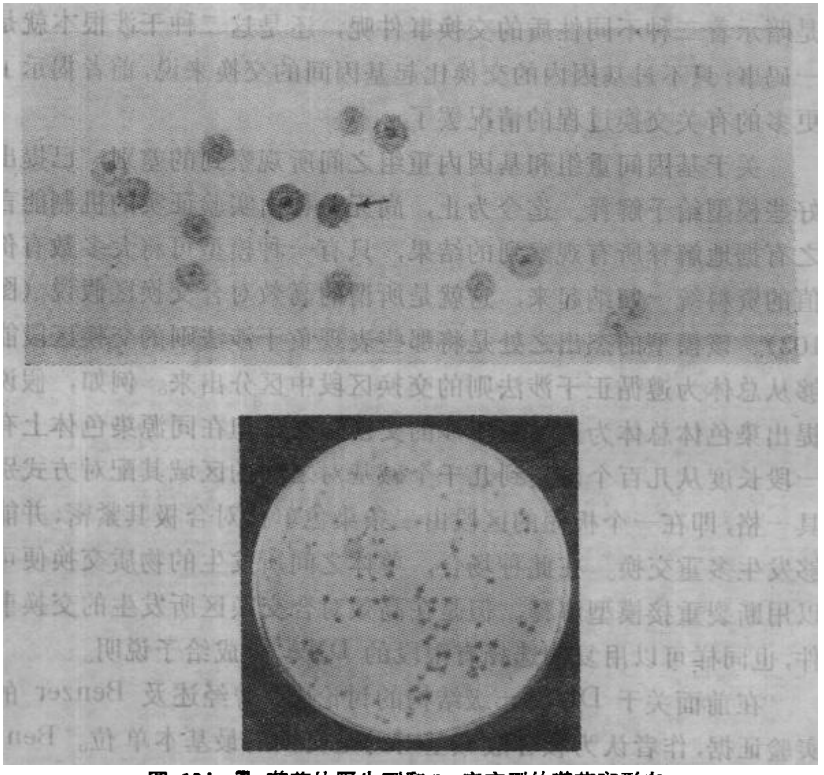


图 104　T_4 噬菌体野生型和 r_{II} 突变型的噬菌斑形态

上面是生长在琼脂培养基表面的大肠杆菌(K12菌株)的光学显微镜照片。噬菌斑(联成一片的透明斑块),来自于T_4噬菌体的感染。野生型 $T_4(r^+)$的噬菌斑较小边缘粗糙,突变型(r)的噬菌斑较大,边缘光滑。r 表示溶菌迅速,噬菌斑亦因溶菌较快而趋于大型。(由 D. P. Snustad赠送。)

细胞时均可表现出生理上的互补效应,尽管并未发生重组(类似于 $AaBb$ 杂合子),但A突变型不能同A突变型互补,B突变型亦不能同B突变型互补。这一规则是采用A和A,A和B,B和B的突变型组合混合感染,并通过典型的顺反子测验判别A与B是否为等位基因的方法加以确立的,在这里正效应可按形成的野生型噬菌斑作为指标。为了确认正效应是来自于互补作用,可将互补测验与重组测验作一对比,对于一切表现互补并看来很象是发生重组的组合,均统计其后代中出现的重组子,根据这两类测验固有的特征加以检测,即可明确无误地区分它们究竟是生理互补,还是重组

表 10 r_{II} 突变型的主体特征

菌　株	野生型 T_4 噬菌体	r_{II} 突变型 T_4 噬菌体
大肠杆菌 B	噬菌斑较小，边缘粗糙	噬菌斑较大，边缘光滑
大肠杆菌 K12λ	与 B 菌株的噬菌斑相同	不能形成噬菌斑或噬菌斑不明显
大肠杆菌 K12	正　常	正　常*

* 大肠杆菌 K12 中原噬菌体 λ 的存在会使 r_{II} 突变型菌株复制受到抑制。

互补。

在诱导成功数千种 r_{II} 突变型的基础上，按照据以精心绘制的遗传图谱（图 90），可将染色体上 r_{II} 这一有限区段里的重组频率与该区核苷酸碱基对的数目加以对比。例如 T_4 DNA 总共含有大约 2×10^5 个核苷酸对，其染色体遗传图长度则接近于 800 个图距单位。因而，1 个图距单位大约相当于 250 碱基对。T_4 噬菌体的染色体 r_{II} 区段将近 5 个图距单位长，亦即约为 1250 核苷酸对，其中 *B* 基因为 3 个图距单位长，相应于 750 碱基对，而 *A* 基因为 2 个图距单位长，即约略等于 500 个碱基对。（在 *A* 基因和 *B* 基因区段现已获得多种缺失突变型，有关遗传图中缺失区段的运用可参见图 105。）

r_{II} 区段内已有 300 个以上的突变型位点被成功地加以鉴别，根据总数为 1250 核苷酸对推算的图距长度，所鉴定的突变型其所占的图距长度也恰好在预期的范围之内。如前所述的每一图距单位平均为 250 核苷酸对，而这里所观察到的最小图距单位（重组值）为 0.01，由之估算这意味着基本单位（指碱基）的重组可发生在相邻的 2.5 个核苷酸之间。另外读者还可以从遗传图中看出，*A* 基因和 *B* 基因中的某些区域似乎较之其他区域更易发生突变，但是，如后面我们还要讨论的，这一问题迄今仍未搞清：突变频率的高低究竟是受 DNA 分子本身性质的限定呢，或是受其他各种因素的制约？例如由 DNA 编码的蛋白质在功能上显不出 DNA 中发生的变化等等。

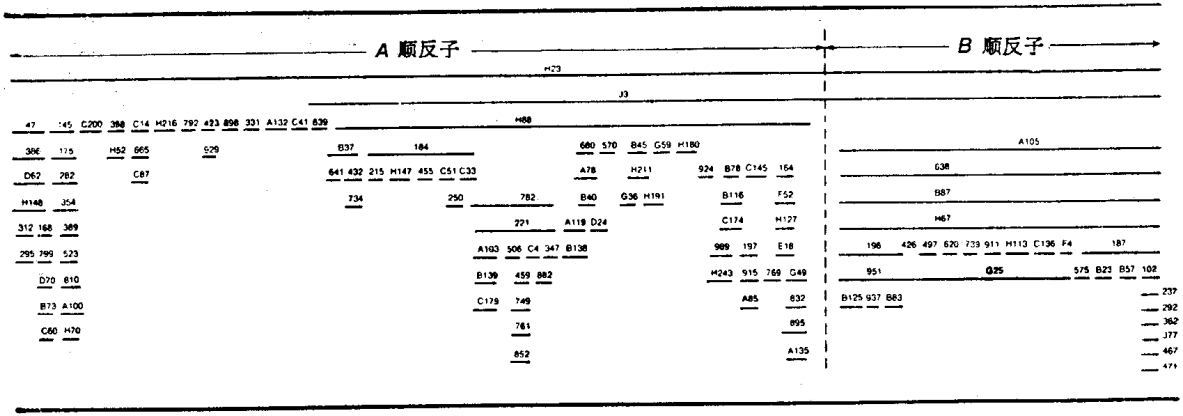


图 105　T_4 染色体 r_{II} 区段缺失突变示意图

T_4 噬菌体 r_{II} 区段有 10%的自发突变是缺失突变，这些突变相互交迭，遍布于整个A基因和B基因区域。在实践上对新突变进行定位时，首先是使用新突变型与作为"标志"的缺失突变型杂交，如果这样的杂交组合不产生重组后代，则新突变的位置即可大体标定在该缺失突变型的相应区域，一旦新突变的大致位置确立之后，进一步可将之与邻接区域的单个点突变的突变型杂交，统计这类杂交的重组频率，可以定出每个新突变的精确位置。(引自 Benzer, 1959。)

Yanofsky 及其同事采用另一不同的系统，即大肠杆菌中控制色氨酸合成酶形成的基因进行实验，亦同样能证明在 DNA 中碱基对之间所发生的遗传学重组。这方面的工作我们打算紧接着遗传密码一节之后讨论，有关 DNA 分子碱基对和蛋白质分子的氨基酸顺序之间关系的探讨亦将在该节作充分展开。但无论如何，根据上面的讨论已经完全可以说，人们确实看到了发生于由每三个核苷酸碱基对组成的碱基组内的重组事件。而且已知每个碱基组正好对应着多肽链中的一个特定的氨基酸。从这些研究来看，似乎重组可以发生在任意二个核苷酸对之间。至于交换(重组)频率的不同倒底是起因于不同的 DNA 中不同区段的差异呢（这方面一无所知)，还是由于存在着某种调节交换频率的控制机制？仍然不得而知。

总而言之，遗传学重组乃是包括二条染色体的同源区域之间发生的染色体片断的物质交换。看来，与交换事件一起，象是这里亦发生有类似于 DNA 的修补合成过程，有充分的理由设想，原核生物中所出现的转录、转化以及染色体交换等等，同上述过程恐怕不无关系。尽管有关 DNA 分子特性的详尽知识对于现行的重组模型已经给予了有力支持，不过就遗传学重组现象来说仍然无法指望能够在分子水平上得到详细阐明。

参 考 文 献

Campbell, A. M.: *Episomes*, Harper & Row, New York, 1969.

Drake, J. W.: *The Molecular Basis of Mutation*, Holden-Day, San Francisco, 1970.

Fincham, J. R. S.: *Genetic Complementation*, Benjamin, New York, 1966.

Ingram, V. M.: *The Biosynthesis of Macromolecules*, Benjamin, New York, 1965.

Kornberg, A.: *DNA Synthesis*, Freeman, San Francisco, 1974.

Stahl, F. W.: *The Mechanics of Inheritance*, 2d ed., Prentice-Hall, Englewood Cliffs, N.J., 1969.

Whitehouse, H. L. R.: *The Mechanism of Heredity*, 2d ed., St. Martins, London, 1972.

Wilson, H. R.: *Diffraction of X-Rays by Proteins, Nucleic Acids and Viruses*, St. Martins Press, New York, 1966.

第四章 RNA：转录与遗传信息的转译

转　录

生物信息的表达是一个过程，而不是整体等于各部分加和的静态集合。在这个过程的开头，我们观察到以 DNA 复制的形式存储信息的现象。复制和存储信息是 DNA 的二大基本功能，DNA 另一重要功能是把信息转移到另一类称做RNA的信息大分子上。信息从 DNA 转移到 RNA 的现象称为转录。转录是以合成 RNA 的方式完成的。和上一代 DNA 分子与下一代 DNA 分子之间的以互补法则作为指导的 DNA 合成一样，RNA 的合成过程使得 RNA 转录本中保持了亲代 DNA 分子的碱基顺序。

信息从 DNA 转移到 RNA 的方式和信息从 DNA 转移到 DNA 的方式大致相同。在这二种场合，DNA 分子都作为模板，并且一般说来，这二种过程的化学反应也是相同的。然而两者的结果和特性却不一样。从严格的信息转移的观点来看，DNA 复制是一种整体合成现象，DNA 复制一旦在细胞中开始，即便从头至尾复制了整个基因组，而且整个复制过程是与细胞分裂或病毒繁殖同步进行的。RNA 的合成则按照另一种模式。信息从 DNA 转移到 RNA 是零打碎敲的，因此基因组内不同基因的转录各自独立。每个基因或基因小群是产生一个或一个以上自身 RNA 拷贝的“单位”。每个单位对某个独立的调节信号各自作出反应。

有关于信息流的总的看法，一般认为信息系从DNA通过RNA而最终到达蛋白质。这意味着信息从受 DNA 复制单位的集中控制过渡到由小的 RNA 复制本的独立控制，也就是说，过渡为一种分散的效应。一个完整的基因组包含有成千上万个基因，而 DNA

的合成方式则保证了细胞分裂期间每个子细胞能得到一套完整的基因组。分配在细胞内的每个基因，随后在个体发育过程中又通过 RNA 转录产生一个个转录本。每一个转录本又指导着为维持

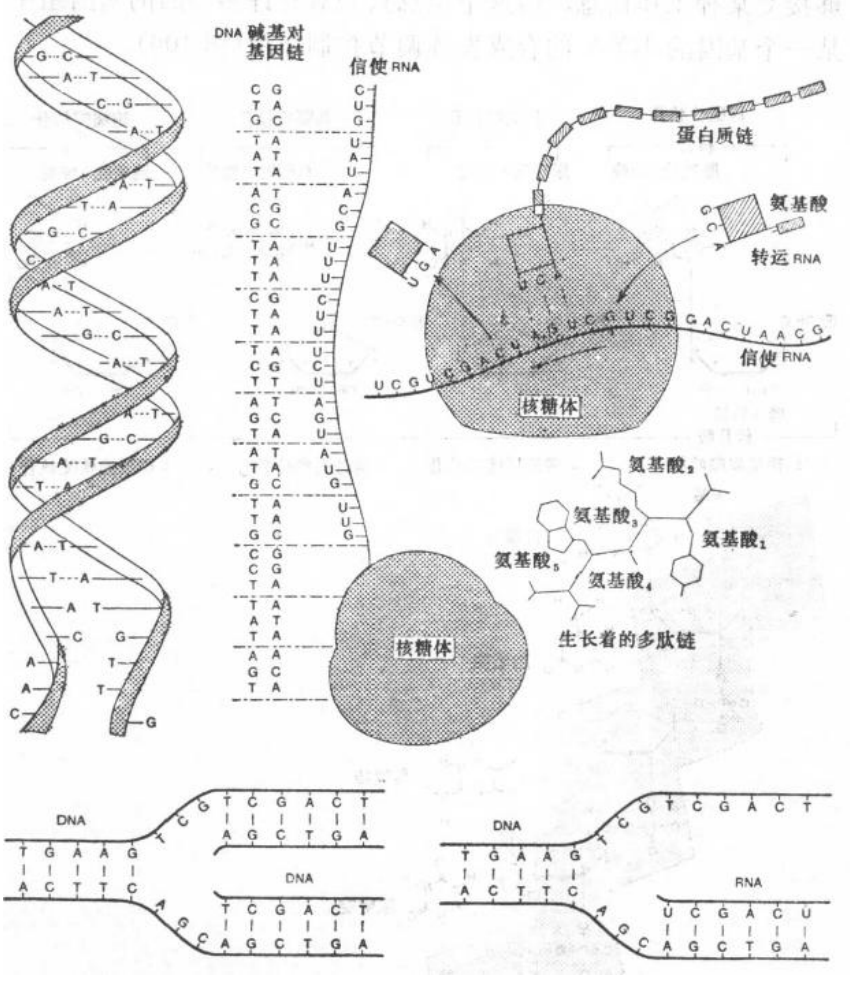
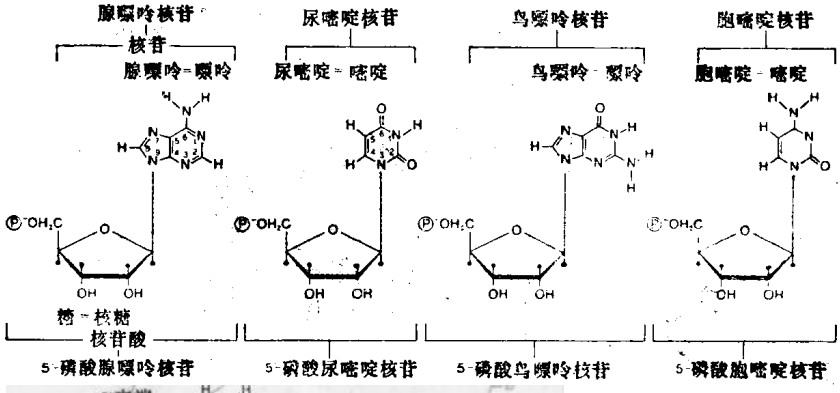


图 106　遗传信息流传的概况

DNA的二个主要功能是复制和转录。复制导致二个新的 DNA 分子的合成。亲代分子的每条链都作为指导子代分子合成的模板。转录产生一个 RNA 分子，DNA基因链作为RNA的模板。RNA分子把来自基因链的信息转移到蛋白质，RNA 的碱基顺序转译成蛋白质分子上的氨基酸顺序。(引自 Crick, 1966, 与 Dickerson及 Geis, 1969。)

细胞或有机体生命活动所需要的千万种蛋白质中的某一种蛋白质的合成。在细胞和有机体的生长和发育期间，每一种蛋白质的合成都有各自最适合的时间和地点，激发 RNA 合成的控制机构能够接受某种工作信息，而这个信息只是对有许多基因的基因组中某一个基因的 RNA 的合成发挥调节控制作用(图 106)。



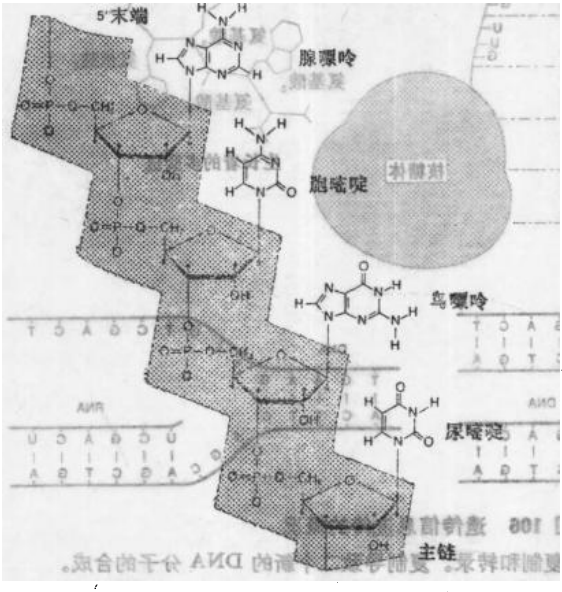


图 107 RNA 分子

构成RNA分子的核苷酸和核糖与DNA分子相似，不同的只是尿嘧啶取代了胸腺嘧啶以及核糖替代了脱氧核糖。图示一小段多聚核糖核酸(RNA)分子。

RNA 分子的一般结构和组成在三个方面与 DNA 分子不同(图 107): (1) RNA 分子含有核糖而不是脱氧核糖。RNA 是由一系列核苷酸亚单位，由核糖和磷酸间 3′→5′ 磷酸二酯键连成的链状结构。(2) RNA 分子一般是单链而不是双链。但作为例外，双链 RNA 在某些 RNA 病毒(如呼肠孤病毒)中亦有发现。某些单链分子以自身折迭的方式在单链的不同位置的碱基对之间形成氢键 (表 11)。由于单链分子自身折迭，RNA 分子中会出现双链部分和单链部分，从而形成某种形式的二级结构。二级结构的专一性一般要高于直线单链分子。(3)作为规律,对应于 DNA 胸腺嘧啶的位置,在 RNA 分子上都是用尿嘧啶代替的。如图 46 所示，胸腺嘧啶和尿嘧啶的差别仅在 5 位碳原子上。氢原子取代了胸腺嘧啶的甲基就成为尿嘧啶。然而，在某些 DNA 分子中有时也含有尿嘧啶。

表 11 RNA 与 DNA 构型比较

构　型	DNA 实　例	RNA 实　例
单链,线状	SV40 动物病毒	Qβ 和 f_2 噬菌体
双链,线状	大多数真核生物	至少某些动物肿瘤病毒是如此(具有部分双链)
双链,环状	线粒体、细菌	呼肠孤病毒(Reovirus)
单链,环状(以双链形式复制)	φX174噬菌体,M13	

接受基因指令产生一个或一个以上自身转录本这一事实，我们可以预期 RNA 分子的长度理应各不相同，而且 RNA 分子的数值必定比 DNA 分子为小。此外,RNA 与 DNA 的化学性质和结构很不相同。必须牢记二者在功能上有着本质性差异。DNA 担负着存储、复制和转录信息的机能,而 RNA 则是把信息转译为蛋白质。转录机制留待以后叙述，这里只先讲明转译是一个把遗传信息从 RNA 转录本的核苷酸顺序转移到蛋白质氨基酸顺序的过程。

1956 年 Ochoa 首次发现了 RNA 的受控合成。他描述了多聚核苷酸磷酸化酶能够在没有模板的情况下把核苷酸连接到正在生长着的多聚核苷酸链上，从而合成没有专一顺序的多聚核苷酸长链。以二磷酸核苷为底物，多聚核苷酸磷酸化酶催化如下反应：

$$n\text{XDP} \xrightarrow{Mg^{2+}} (\text{XMP})_n + n\text{P}i, \qquad \text{X}=\text{任何碱基}$$

控制该系统产生 RNA 产物成分的唯一方法是控制底物分子的数量和种类。如果所有的 XDP 都是二磷酸尿嘧啶核苷，那末 $(XMP)_n$ 将是多聚尿嘧啶核苷酸。如果 XDP 由 10 个二磷酸尿苷、5 个二磷酸腺苷和 1 个二磷酸鸟苷所组成，那么合成的 $(XMP)_n$ 的核苷酸比例也将与之相同，但是没有专一的顺序。多聚核苷酸磷酸化酶不能复制模板的信息，因而在这个酶发现以后的一段时间内，仍不了解 RNA 容纳的信息和信息从 DNA 到蛋白质的转移机制。

在 Ochoa 报道无专门顺序的 RNA 离体合成之后不久，人们的注意力即开始转向 RNA 病毒和它们的繁殖方式。RNA 病毒不包含 DNA，某些病毒的 RNA 是单链，而另一些是双链。RNA 病毒种类繁多，反映了形形色色的 RNA 复制机制。然而，RNA 合成具有共同的特征。

所有的病毒在宿主细胞中增殖，因此所有的病毒 RNA 都是在宿主细胞内进行合成。大多数病毒复制 RNA 的形式与 DNA 病毒复制 DNA 的形式相同。这些病毒的遗传信息储存在 RNA 内，新 RNA 产生于模板 RNA。另一类病毒，RNA 染色体只有在 DNA 中间体合成之后才复制。中间体一旦形成，信息从 DNA→RNA→蛋白质的正规转移顺序也就据以建立。在初始阶段，DNA 中间体的意义曾遭到忽视。然而这是一个鼓舞人心的发现，信息居然会从 RNA 染色体转录成 DNA 转录本。根据信息可以逆向流传的事实，很快有人提出，有可能在体外通过 RNA 转录本制造出基因。换言之，任何来源的 RNA，加上合成 DNA 所需要的底物及反向转录酶——识别 RNA 模板合成 DNA 的酶——便有可

能人工合成"基因"，随后"基因"在 RNA 多聚酶的作用下，又进而作为模板再度合成 RNA 分子。

从认识 RNA 合成过程的观点来看，不产生 DNA 中间体的病毒显得更为重要。Sol Spiegelman 和他的同事曾详尽地研究了这些病毒中一种叫做 Qβ 的病毒。Qβ 是一种很小的病毒，它的染色体是仅仅包含几个基因的单链 RNA 分子。其中一个基因编码的蛋白质，组成病毒的外壳或外壁(称为外壳蛋白质)；另一个基因编码的多肽是一种酶——RNA 复制酶，它在宿主细胞中催化 RNA 复制。

Qβ 病毒感染大肠杆菌，感染之后引起细菌溶菌的步骤完全类似，但是和 DNA 病毒溶菌的步骤则不尽相同。继觉察病毒染色体插入宿主染色体之后立刻发现的一个不同之处是 DNA 病毒染色体的复制几乎马上开始，而 Qβ 染色体在感染之后，第一步却是转译。即 RNA 病毒的一个基因的信息(以合成蛋白质的方式)转译为组成 RNA 复制酶四种多肽的一种。也就是说，没有复制酶的存在，RNA 染色体不能复制。有趣的是复制酶另外的三种多肽由宿主细胞的染色体编码提供，仅仅第四种多肽，即 β 肽链由 Qβ 染色体(为 RNA 染色体——译注)编码。复制酶一旦装配完成，以Qβ染色体为模板，合成新的Qβ染色体即行开始(图 108)。

这种识别 RNA 模板的复制酶，被很恰当地称为依赖于 RNA 的 RNA 多聚酶，它类似于依赖于 DNA 的 DNA 多聚酶，是由 4 条不同肽链组成的又大又复杂的蛋白质。这里饶有趣味的事实是只有来自两个不同有机体的基因(一个来自 Qβ 病毒，三个来自大肠杆菌)处在同一个细胞中复制酶才能产生。

这里存在着深入探讨宿主(大肠杆菌)和寄生物(Qβ 病毒)相互关系的足够理由，因为就 Qβ 病毒的生存而言，它的染色体必须在宿主细胞中复制。而复制一旦发生，宿主细胞就必定死亡。从宿主细胞携带为寄生物染色体复制必需的大多数遗传信息来看，为什么宿主细胞居然会以自寻绝路的方式向寄生物提供信息呢？而且当宿主细胞不接触寄生物时，这些基因对细胞本身的生存又

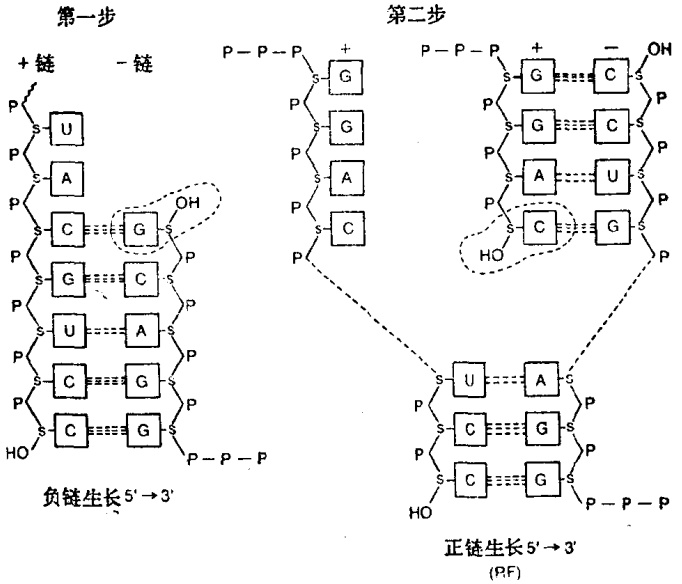


图 108　Qβ 噬菌体RNA 的复制

Qβ 染色体(RNA)进入细菌细胞，其中一个基因借助宿主细胞的蛋白质合成系统产生RNA复制酶的β肽链。在具有功能活性的复制酶作用下，RNA染色体复制周期活动的第一个步骤是合成 Qβ RNA 负链。负链通过与正链上互补的碱基配对形成双螺旋 RNA 分子。这种双螺旋RNA 称为复制型 RNA (RF)。如图所示，RF RNA以 5′到 3′方向优先产生正链(第二个步骤)。(引自 Bishop等，1967。)

有什么潜在的价值呢？无疑，许多细菌宿主和它们噬菌体寄生物之间可能也存在着这种类型的宿主-寄生物相互关系。有理由设想真核类细胞和致癌病毒的关系恐怕与之也不相上下。

RNA 合成期间，作为新 RNA 合成模板的高分子 RNA 并不产生与自身相同的复制物 (正如 DNA 模板不产生本身的精确复制物一样)。RNA 模板产生的仅是它们的补体。亦即原先的 Qβ 染色体 (一个单链的 RNA 分子) 只产生与它互补的 RNA 子链。然后再由子链作为合成新链的模板，所产生的新链才和亲代分子完全相同 (图 108)。一般习惯把原先的染色体叫做(+)链，而第一次互补产生的补体叫做(－)链。

既然 RNA(+)链和(－)链是互补的，它们便能够构成双螺旋

RNA 分子。事实上，第一次确认的 Qβ RNA 合成产物正是包括(+)链和(−)链的双链分子 (图 108)。这个双链分子称为复制型(RF)Qβ RNA。RF RNA 分子一产生，就作为合成新(+)链的模板。复制酶能够识别(+)链和(−)链，也能够识别 RF RNA，但是在 RF RNA 作模板时，(+)链可以优先合成。

Qβ RNA 的合成并不违反信息流的规则，但是 Qβ 的信息系统能够在没有(病毒)DNA 的情况下正常工作，即通常由 DNA 执行的功能在 Qβ 中依然存在，Qβ RNA一身二任，既担负自身复制又同时执行把信息转译成蛋白质的功能。

第二大类病毒叫做致癌(oncogenic)病毒，能诱发某些种类的哺乳动物产生白血病和肉瘤，它们在繁殖开始时则必定要产生 DNA 中间体。致癌病毒和 Qβ 噬菌体一样，为单链 RNA染色体，所不同的是致癌病毒携带着反向转录酶的信息密码子，或者更正确地说，是编码依赖于 RNA 的 DNA 多聚酶的基因。致癌病毒的宿主细胞在依赖于 RNA 的 DNA 多聚酶合成之后，即被"指令"进行下列程序的反应：

$$\text{RNA}(+)\xrightarrow{1}\frac{\text{RNA}(+)}{\text{DNA}(-)}\xrightarrow{2}\frac{\text{DNA}(-)}{\text{DNA}(+)}\xrightarrow{3}\text{转录本}$$

这个反应序列的第一步反应由依赖于 RNA 的 DNA 多聚酶催化，第二步由依赖于 DNA 的 DNA 多聚酶催化，第三步由依赖于 DNA 的 RNA 多聚酶催化。其中催化第一步反应的逆转录酶是由 Howard Temin 和 David Baltimore 首先记载的，这个酶的发现曾经轰动一时，并已被用来制造试管基因。而最后一步反应的酶(依赖于 DNA 的 RNA 多聚酶)则早在 RNA 合成过程中已有所了解。

合成 DNA 和 RNA 分别以脱氧核糖三磷酸核苷或核糖三磷酸核苷为底物，二个系统都需要二价金属离子 Mg^{2+} 和现成模板，两种链的生长方向相同，都是从 5′端到 3′端 (表 12)。关于二个系统的差别仍留有一些尚未解决的问题。例如，依赖于 DNA 的 DNA 多聚酶必须同时识别(+)链和(−)链，并且以相等的数量进

表 12 RNA 多聚酶和 DNA 多聚酶的特征*

	RNA 多聚酶	DNA 多聚酶 I
相同点:		
DNA 模板方向	相 同	相 同
对四种三磷酸核苷的需要	需 要	需 要
对二价阳离子的需要	需 要	需 要
5′到3′合成,反向平行	相 同	相 同
不同点:		
功 能	转 录	复 制
分子量(道尔顿)	480,000	109,000
亚基数	5	1
起始点	有	无
末端终点	有	无
对残基顺序和功能链的识别	能	不 能
以完整的双链为模板	是	不 是
产 物	短,游离	长,双链
修补功能	无	有
离体系统中核苷酸周转速率	~50	~10
与机体内核苷酸周转中的速率比较	相 同	~1%

* 引自 Kornberg, 1974。

行复制。虽然亲代 DNA 两条链的信息内容不同,一条是信息链,另一条是它的互补链,但是在 DNA 合成时,两条链同等对待,同时复制。结果便产生与亲代相同的双螺旋子代 DNA 分子。可是,在 RNA 转录时,依赖于 DNA 的 RNA 多聚酶必须仅仅识别基因链,也就是说,必须分辨 DNA 两条链的差别,并在特定的位点开始,只复制其中的一条。

研究 RNA 合成的结果表明,在离体条件下,RNA 多聚酶通常只复制模板 DNA 的一条链。于是便产生了这样的问题:功能

基因只存在于一条链上，还是存在于二条链上？用于进行合成分析的离体系统的状态有否破坏或模糊了机体体内合成的真正性质？无论在离体，还是在体内，我们都不知道是什么因素决定着RNA合成的起点和终点，转录本的长度和每个基因的拷贝数。

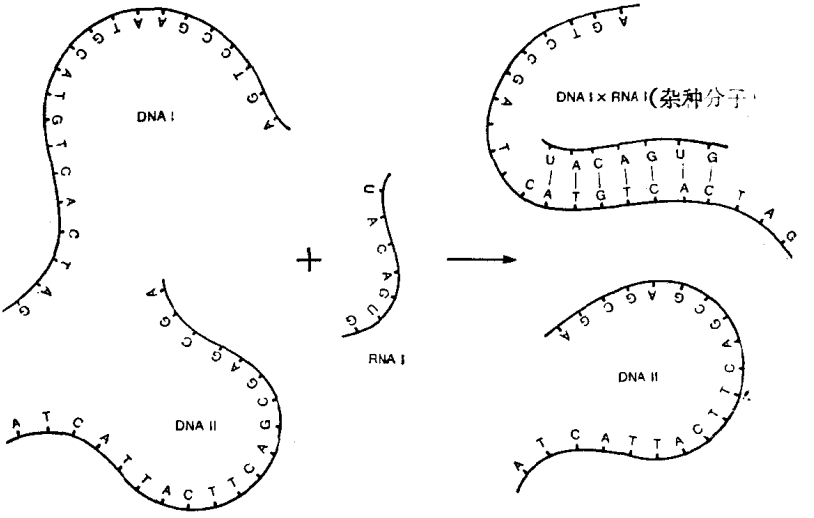
多聚酶分子的性质提供了解答这些问题的一些线索。枯草杆菌(*Bacillis subtilis*)的依赖于DNA的RNA多聚酶是一个分子量约为450,000道尔顿并由五条链组成的巨大的蛋白质分子（表13)。这样缔合的巨大蛋白质是不稳定的。因为当酶从它的天然位置移动时，例如从细胞中分离酶分子，各个亚基往往彼此分离。亚基间由连结力较弱的非共价键组合成复合蛋白质，改变离子浓度、温度、pH等会导致各个亚基的解离。要想把分离和提纯酶的介质条件模拟成机体体内的介质条件决非容易。所以往往不容易分离得到天然结构的复合蛋白质。

表 13　RNA 多聚酶的亚基

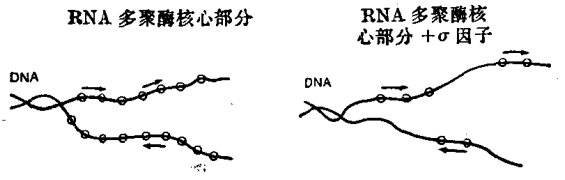
亚　基	分子量（千道尔顿）		每个酶分子的亚基数
	大肠杆菌	枯草杆菌	
α	40	43	2
β	150	150	1
β'	160	150	1
σ	90	55	1
ω	10		?

RNA多聚酶就是一种这样的蛋白质。它带有或不带有σ亚单位的不同性能增加了确定其专一性的困难。在分离RNA多聚酶期间，σ因子很容易丢失，但是σ因子的丧失并不影响转录酶本身的活性，只是使之再也无法辨认RNA合成的专一性起始点。没有σ因子的$\alpha_2\beta\beta'$核心酶蛋白能够转录任何DNA模板，而不表现一丝一毫专一性。在附加σ因子之后，多聚酶只从模板的专门位

点开始合成 RNA（图 109）。这个发现澄清了一个令人困惑已久的问题。既然早期在离体系统中使用的 RNA 多聚酶没有 σ 因子，那么所得到的结果便不能作为体内 RNA 合成的可靠依据。倘若使用全酶，这时 RNA 多聚酶仅转录每个基因的一条 DNA 链，而



(a)



(b)

图 109 RNA 合成起始的专一性

(a) 处于热变性状态DNA的双链解旋，有亲缘关系的RNA仅与其中有互补碱基部分的(DNA)链形成杂种分子。通过密度离心能够分离出混和在单链DNA 以及RNA中的RNA-DNA杂种分子。解开杂种分子得到的 DNA（基因链），可提供进一步研究。(引自 Spiegelman, 1964。)

(b) 使用带有σ因子的RNA多聚酶与不带 σ 因子的RNA多聚酶来研究RNA的合成方式，发现合成起始的专一性系依赖于σ因子。离心分离热变性的双链，分别再用专门的 mRNA 探测每条DNA单链，即可分辨出基因链。一般要确认与指定的mRNA相应的基因链时均使用这种方法。

且 RNA 合成亦被限定从专门的位点开始直到产生一个分子为止(图 109)。

利用带有或不带有 σ 因子的 RNA 多聚酶，能够在一定程度上分析细菌染色体的局部特性。例如，细菌染色体的两条链都具有起始位点，但这些位点从不处于对应的位置上，也就是说，二条链都有基因存在，但某一个基因，只存在于其中的一条链上。RNA 多聚酶 σ 的功能是找出每个基因的基因链，然后按 5′到 3′方向进行复制。从染色体或 DNA 的观点来看，极性和转录可朝二个方向伸展，但从基因的观点来看，极性和转录只能单向伸展（图 110)。

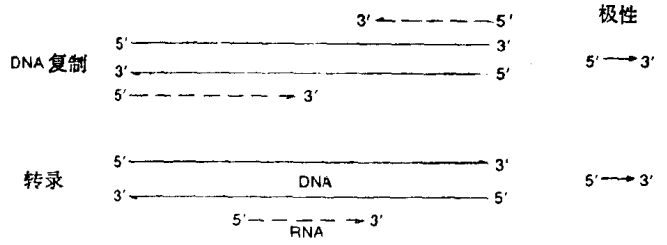


图 110　转录的极性

DNA和RNA的合成、复制和转录的极化方向相同(5′到3′)，DNA复制时，两条链都作为模板(上)，但在转录时，只有基因链作为模板(下)。虽然基因存在于两条链上，但不在对应位置。

因为 RNA 合成还要依靠另外一些因子，它们是一些不由噬菌体染色体编码的分子，所以 RNA 合成远比图 108 所示的一般模式来得复杂。事实上，业已证明那些从未接触过 Qβ 噬菌体的宿主细胞中早已存在着这些因子。虽然目前对这些分子本身或它们的作用机制还一无所知(图 111)，但这些因子在决定 RNA 合成的特征方面肯定起着重要的作用。例如，有一个分子量为 49,000 道尔顿，对热敏感和能被蛋白质降解酶(水解蛋白酶)消化的因子，和 Qβ RNA 结合后，能修饰它的形状。在复制之前，Qβ RNA 在 3′末端有一个腺嘌呤核苷酸。据认为该因子修饰 RNA（原文为 DNA——译注)的方式即是切除末端核苷酸（腺嘌呤核苷酸由于

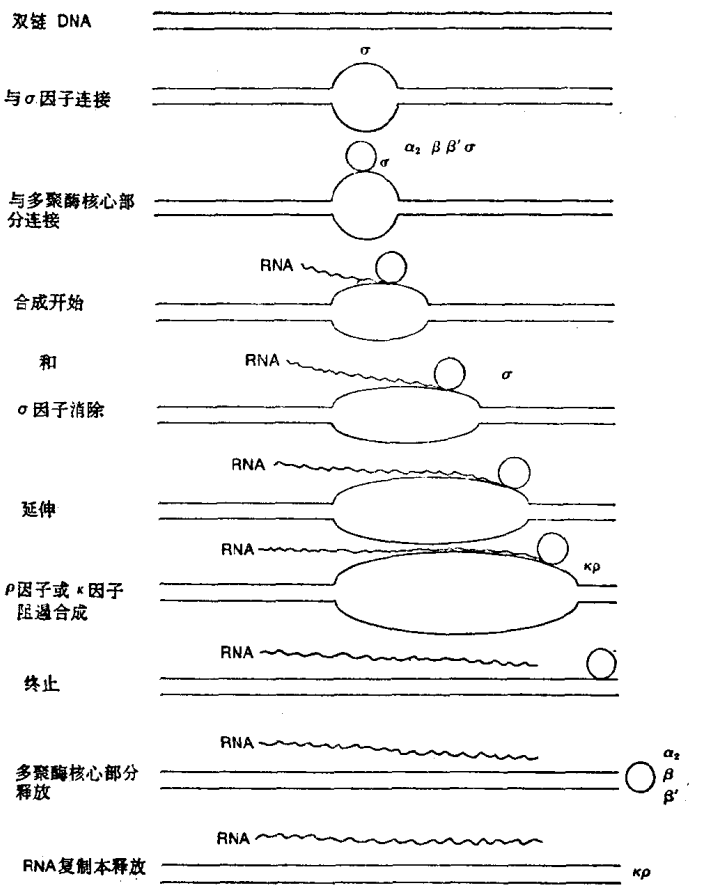


图 111 影响转录的因子

转录的识别步骤:

1. RNA多聚酶和DNA模板连接
2. 链开始生长
3. 链的延伸
4. 链停止生长和RNA多聚酶释放。

因子:

σ: 为多聚酶辨认模板的起始位点所必需，σ因子在和多聚酶分离时不活泼。

ω: 存在于某些大肠杆菌多聚酶制剂中，但有些酶制剂不具备ω因子。其功能不详，或许为链的延伸所必需。

ρ: 功能不详，或许与链的终止合成有关。

κ: 功能不详，或许与链的终止合成有关。

某种尚不了解的机制，在 RNA 合成之后，仍被放回原处)。后来得知，这个因子乃是一种外切核酸酶。

另一个因子其分子量为 80,000 道尔顿，对热反应相对比较稳定，且与 Qβ RNA 连接的亲和力很高，对它的结构和功能人们迄今一无可知。很可能这些因子以及另外的因子对体内 RNA 天然合成十分重要。

因为合成抑制剂中包括一些抗生素，所以研究 RNA 合成的机理在实践上极为重要。那些能控制和预防细菌对人体以及能控制和预防对家畜感染的药物一般地都被泛称为抗生素。某些抗生素能抑制原核生物的转录，但对于真核生物细胞中的转录过程没有妨碍；另一些抗生素的作用则与此恰恰相反。关于大多数抗生素的作用机理认识还很肤浅。但是有一种仅对原核生物有效的叫做利福霉素的抗生素，据认为因其和 RNA 多聚酶的一个亚基相连接，所以能够阻止多聚酶的启动。一般而言，能够阻止原核生物转录的抗生素往往也能阻止线粒体和叶绿体的转录过程。

放线菌素D是这一总的分类原则的一个例外。它能阻止很多类型细胞的转录。它的作用似乎是嵌在 DNA 模板鸟嘌呤残基的旁边，以阻止 RNA 越过嵌入点继续合成。放线菌素D嵌入 DNA 的识别信号，据认为是$\frac{GC}{CG}$。生长着的 RNA 链一般越过嵌入点最初四个 DNA 核苷酸对就马上停止前进。曾观察到在以人工合成的缺乏鸟嘌呤的 DNA 作模板时，放线菌素D对 RNA 合成的抑制作用很不灵敏，才由此作出了这个明确的结论。

真核生物细胞核内转录的复杂程度远远超过原核生物。这里仅举一例：三类RNA 分子各由一种专门的 RNA 多聚酶催化。一种集中在核仁内的 RNA 多聚酶专门催化核糖体 RNA (rRNA)的合成，核糖体 RNA 是细胞器核糖体结构的组成部分。另外还有二种位于核仁之外但存在于胞核中的多聚酶，一种专门催化合成信使 RNA(mRNA)，另一种专门催化合成转运 RNA(tRNA)(图 112)，彼此在特性上是截然不同的。

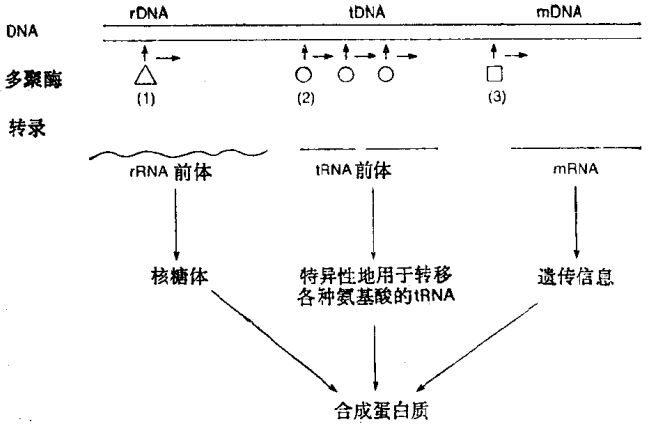


图 112　真核生物的三种RNA多聚酶和三种RNA

图示三种RNA：rRNA，tRNA和mRNA的功能。在真核生物中，每种RNA由特殊的RNA多聚酶合成。一种位于核仁中，二种位于胞核内。在转录过程中，没有发现这些RNA的差别，但在转录之后，发现它们在大小、组成、形状和功能方面彼此很不相同。三种RNA在促进转译方面则相互协作，以合成蛋白质的方法把遗传信息转移到蛋白质分子上。

转　　译

信息从 RNA 转录本转移到肽链称为转译。现在一致公认传递过程的生物学机理就是蛋白质合成，转译和蛋白质合成经常被当作同义词使用。但是，按照科学的含义，转译乃是信息流动的一个步骤，蛋白质合成还包含有生物学上的连接机理。

蛋白质合成过程伴随有大量的协同有序的生物化学反应。就这点来说，蛋白质合成比 DNA 和 RNA 合成更加复杂。蛋白质合成所需要的三种 RNA 的生成机制相同，但是它们在生成和进一步加工之后，在蛋白质合成过程中却各有其专门的功能。在着手研究蛋白质合成很长一段时期之后，方才最终鉴别和确定出这些 RNA 的特异性功能。

五十年代 Paul Zamecnik, Malon Hoagland, Fritz Lipmann 等人曾勾划了蛋白质合成的大致轮廓。可见早在遗传信息从 DNA 到蛋白质的传递情况完全搞清之前这一工作就已经开始

了。但直到五十年代后期，所得到的有关 DNA 复制、RNA 合成和蛋白质合成方面的线索逐渐汇集到一起。从此以后，经历上述各个步骤的遗传信息流传才多少被赋与某种含义。

揭示蛋白质合成规律的首次实验是在细胞质而不是在细胞核或染色体中进行的。如果把放射性同位素引入旺盛生长的真核生物细胞，不久即可观察到放射性主要集中在叫做内质网的细胞质膜系统上。进一步检查内质网时又发现所显示的放射性是围绕着核糖体的，确实，聚集放射性的核糖体正是生长着的肽链附着的场所（图 113）。

第二个实验是关于 tRNA 在蛋白质合成中的作用。tRNA 是

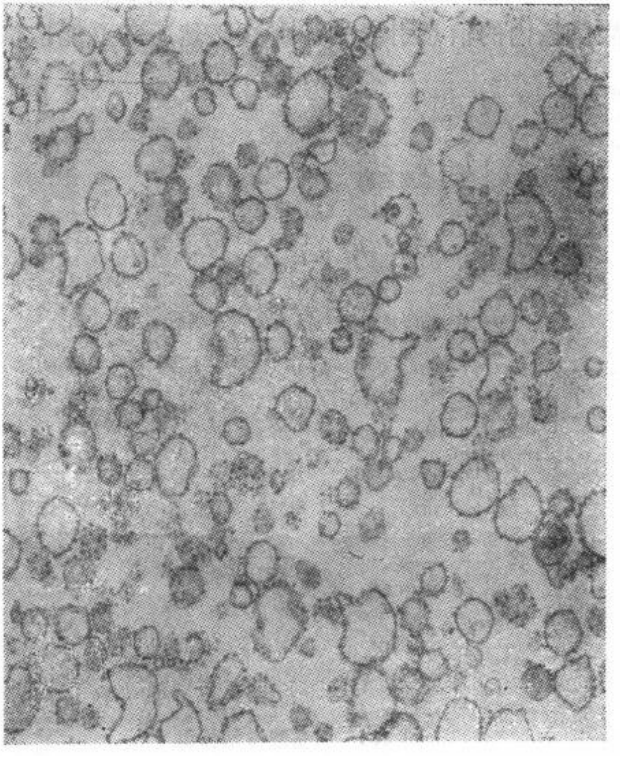


图 113　微粒体膜

微粒体膜组成称为内质网的细胞膜体系，核糖体（图中小黑点）附着在膜上. 细胞中大部分蛋白质在此合成。

最小的核酸分子，它的长度在 75 到 85 个核苷酸之间。因为体积小，所以 tRNA 在使用沉降法分离大分子化合物的条件下，仍然可以溶解而不沉淀。事实上，起初它被称为可溶性 RNA，或 sRNA。tRNA 分子的功能是把氨基酸运送到合成蛋白质的特定地点。一个 tRNA 分子运送一个氨基酸。每个氨基酸在肽链上的着生位置按照模板 RNA 分子提供的信息来确定。氨基酸开始附着于 tRNA 分子上的机理称为氨基酸激活作用，起始附着由每个氨基酸-tRNA 组合的专一性酶催化（图 114）。氨基酸附着到 tRNA 上以后，就被传递到生长着的肽链上。

甚至早在鉴别 tRNA-氨基酸复合物是信息流程上的重要步骤之前，Crick 就提出了一个假设，即如果肽链的生长是由包含在核苷酸模板内的信息引导的，那么必然还需要能按信息从而把

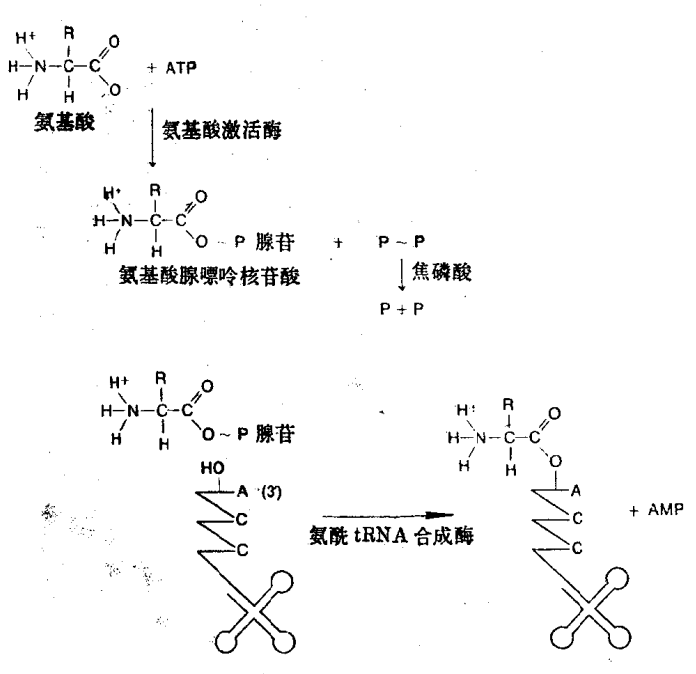


图 114　氨基酸激活作用

氨基酸由ATP激活，与 tRNA 分子 3′末端的腺嘌呤残基形成化学键。氨基酸和tRNA分子的专一性结合乃是某种催化形成tRNA-氨基酸复合物的酶的功能（这些酶称为氨酰tRNA合成酶）。

每个氨基酸插到生长着的肽链合适位置上的接合子分子来执行将氨基酸特定排列的职能。现在已经知道，tRNA 分子正是执行这一功能的生物大分子，因此许多科学文献把 tRNA 分子称为接合子。

催化氨基酸与 tRNA 复合的激活反应的酶叫做氨酰-tRNA 合成酶，这些酶催化的总反应为：

$$\text{ATP}+\text{氨基酸}+\text{tRNA}\longrightarrow\text{氨酰-tRNA}+\text{PP}i+\text{AMP}$$

氨基酸结合到生长的肽链上去之前必须预先活化，在开始认识这一点时，只知道每种氨基酸的化学结构是唯一的，但不知道每种氨基酸是否只有一种 tRNA 还是有不同的 tRNA 与之相对应。假定氨酰-tRNA 复合物的氨基酸部分在蛋白质合成期间识别模板，在理论上是行得通的。这个假说在五十年代中期的一段不太长的时期内曾经流行一时。但在 tRNA 分子为氨基酸识别模板提供了一种更好的解释之后，人们对一般氨基酸活化概念的理解也就更加完善和深入了。例如，如果是 tRNA 分子把氨基酸引导到它们在肽链上的插入位置，那么每种氨基酸的氨酰-tRNA 合成酶必须是独一无二的，现在，这一点已由实验证实。

转译是信息从基因到蛋白质过程中的一个步骤。如上面讨论的那样，模板作为初始的信息源，通过规定蛋白质合成期间肽链上氨基酸的排列顺序而起作用。换言之，每个模板都以其专一的核苷酸顺序起着基因信使的功效。历来试图说明这个现象的假说层出不穷，妙趣横生——多到了难以介绍的程度。然而这个现象的正确解释最后则是由 François Jacob 和Jacques Monod 在 1961 年确定的，他们首先指出蛋白质合成时的模板是信使RNA(mRNA)(图 115)。mRNA 分子直接拷贝于基因，由其单链的核苷酸顺序规定着肽链的氨基酸顺序。因此，它们作为模板的机理就在于此。既然 mRNA 分子是单链，所以它的核苷酸能和任何与之互补的核酸(DNA 或 RNA)自由地形成氢键。正是 mRNA 小段核苷酸和 tRNA 之间的这种互补，才保证了蛋白质合成中氨基酸的恰当排列顺序(图 116)。

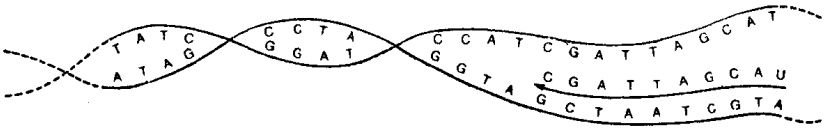


图 115 信使 RNA

双链DNA只有一条链是基因链，或者说是包含信息的链，在RNA多聚酶的酶促作用下，RNA转录本直接拷贝基因链。mRNA分子的寿命一般很短，它们在细胞中的浓度很低，很少超过细胞中RNA含量的百分之十。

氨基酸的激活和附着到 tRNA 上是在一个化学反应中完成的。tRNA 分子对它们附着的氨基酸有专一性，这个专一性由 tRNA 对 mRNA 的识别来体现。但这个专一性不能解释蛋白质的合成机制。如上所述，蛋白质是在核糖体上合成的，核糖体是一种很小的，为真核细胞、原核细胞所共有的细胞器。就它的重量而言，核糖体几乎一半是 RNA，一半是蛋白质。核糖体蛋白质部分在蛋白质合成时作为识别位点，并起催化形成肽键的酶的作用，是蛋白质合成不可缺少的基本因素。核糖体 RNA 的功能还不清楚，它似乎起某种骨架的作用，在适当位置上承受蛋白质。

然而核糖体的总的功能已经了解得相当清楚，核糖体与 mRNA 分子形成复合物。氨酰-tRNA 分子就附着在这个复合物上。核糖体提供催化形成肽键的肽基转移酶和生长的肽链的附着位置。此外，还以某种还不知道的方式确保 mRNA 分子经过肽链的附着位置向前移动，以致于使每个新的氨酰-tRNA 复合物都能恰好不过地与生长着的肽链相连结。每当一个氨基酸与肽链的末端羧基形成肽键以后，携带该氨基酸的 tRNA 分子就从复合物中释放出来，并且再通过激活过程循环使用。氨基酸连接完毕，mRNA-核糖体复合物立即拆除，释放出结构完整的肽链。留下的核糖体以后再附着于另一个 mRNA 分子重新运转。从这个简短的关于蛋白质合成的叙述中，可清楚地看到，合成过程包含几种 RNA，每种 RNA 以其固有的结构上或信息上的特点在合成过程中担任专门的角色。深入研究每一种 RNA 的结构和功能对于进一步理解蛋白质的合成具有十分重要的意义。

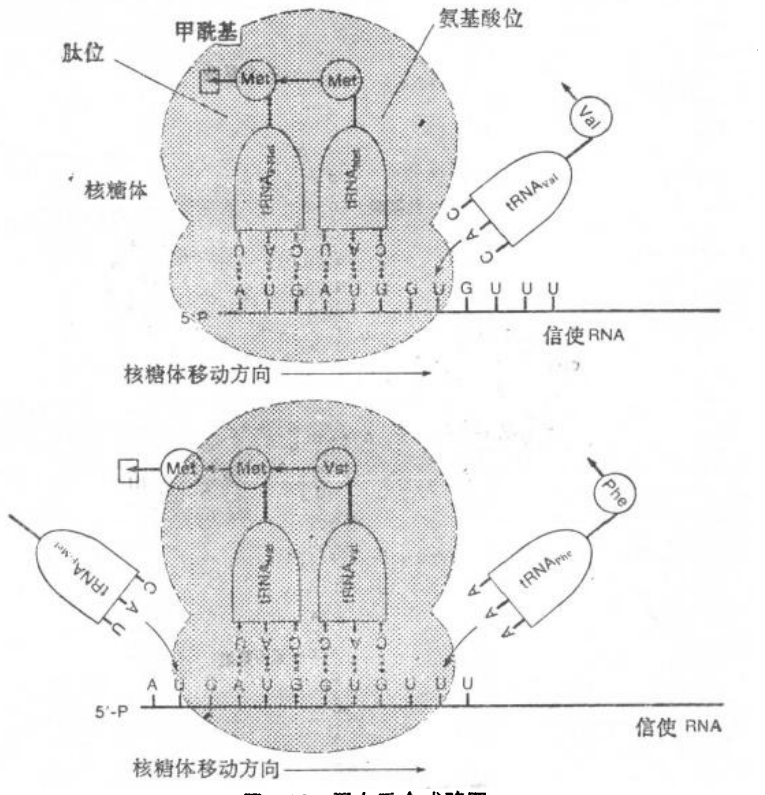
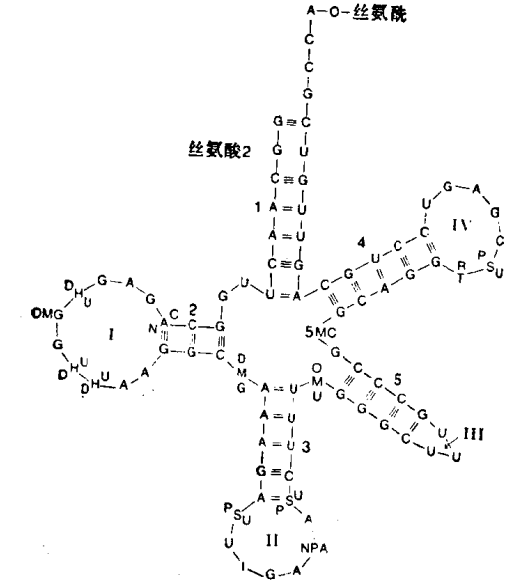


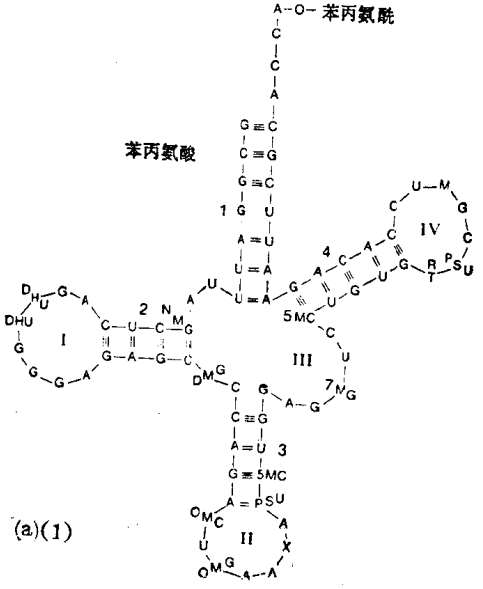
图 116　蛋白质合成略图

蛋白质合成在核糖体上进行，肽链的氨基酸顺序由mRNA分子的核苷酸顺序所决定。氨基酸由tRNA分子携带至肽链的生长位置。核糖体有二个供氨酰-tRNA分子附着的位置，即肽位(P)和氨基酸受位(A)。第一个氨酰-tRNA直接和P位相连，蛋白质合成由此开始，第二个乃至其余的全部 tRNA都是先附着到A位，在A位的氨基酸与P位的氨基酸或肽形成肽键之后，A位的tRNA即移到P位。肽键形成和氨酰-tRNA从A位转移到P位同时发生。肽键形成之后，P位的 tRNA和它所附的氨基酸分离，从复合物中释放。同时第三个氨酰-tRNA占据A位，整个过程反复循环，直到mRNA的信息转译完毕，即一条完整肽链形成为止。(引自 B. F. C. Clark 与 K. A. Marcker, 1968。)

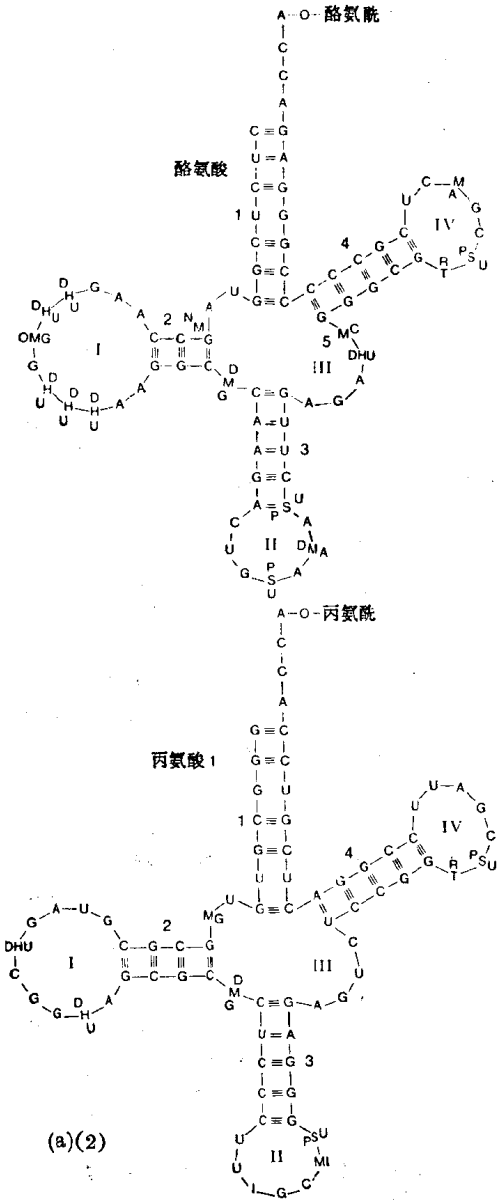
转移 RNA

tRNA 分子是一类迄今已知的最小的核酸分子，它们的核苷酸数在 74～90 之间，总共约占整个细胞 RNA 的 15%。tRNA 分子和其他 RNA 分子的差别在于它含有一些稀有碱基，并且通过直链分子不同部位的碱基相互配对产生双螺旋区域（图 117）。所

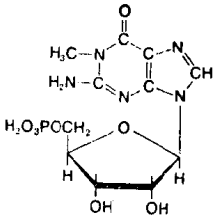




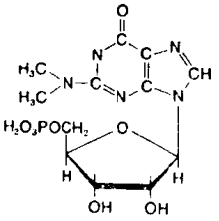
(a)(1)



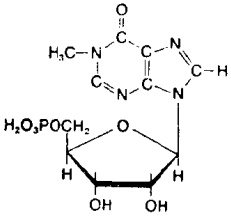
(a)(2)



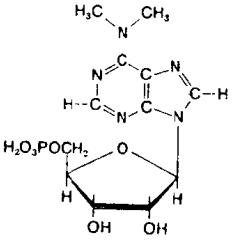
1-甲基鸟嘌呤核苷酸
(MG)



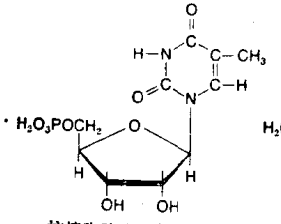
N(2)二甲基鸟嘌呤核苷酸
(NMG)



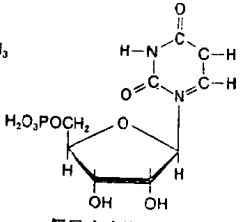
1-甲基次黄嘌呤核苷酸
(MI)



N(6)二甲基腺嘌呤核苷酸
(DMA)

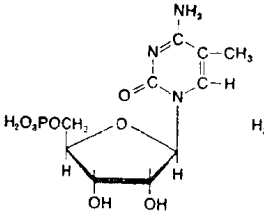


核糖胸腺嘧啶核苷酸
(RT)

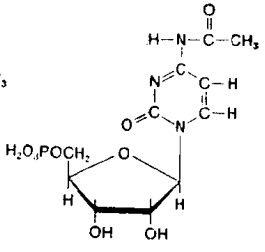


假尿嘧啶核苷酸
(PSU)

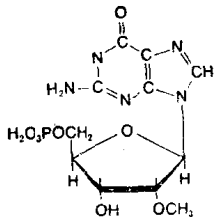
(b)(1)



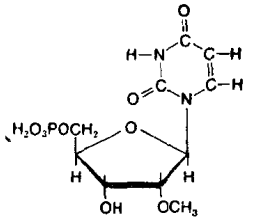
5-甲基胞嘧啶核苷酸
(5MC)



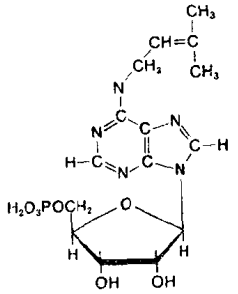
N(6)乙酰胞嘧啶核苷酸
(NAC)



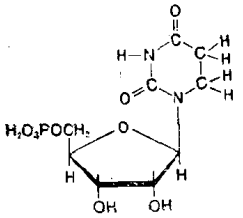
2′-O-甲基鸟嘌呤核苷酸
(OMG)



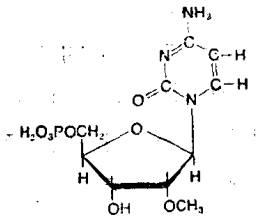
2-O-甲基尿嘧啶核苷酸
(OMU)



N(6)-异戊烯腺嘌呤核苷酸
(NPA)



4,5-二氢尿嘧啶核苷酸
(DHU)



2-O-甲基胞嘧啶核苷酸
(OMC)

(b)(2)

图 117 tRNA 分子

(a) 直线单链分子互补部分之间形成由氢键联系的区域决定tRNA分子的形状。分子的不配对部分形成环。现已发现识别位点均在环状区域之内。如环Ⅱ有识别 mRNA 密码子的反密码子，环Ⅳ或许有核糖体的连接位点，环Ⅰ被看作氨酰合成酶的识别位点。

(b) 转录之后RNA的正常碱基由酶转换或修改为稀有碱基。稀有碱基的形成是在转录之后一种复杂的加工现象，为 tRNA 分子提供专一性。(引自 R. W. Holley, J. Apgar, G. A. Everett, J. T. Madison, M. Marquisse, S. H. Merrill, J. R. Penwick, 与 A. Zamis, Science, 147:1462, 1965; H. G. von Zachau, D.Dutting, 与 H.Feldman Angew Chem, 1966; J. T. Madison, G. A. Everett 与 H. Kung, Cold Spring Harbor Symp Quant Biol, 31, 1966。)

有 tRNA 分子的结构具有一定的共同构型。例如，分子的 3′末端的核苷酸都是不配对的，而且最后三个核苷酸一定是 CCA。在活化反应中所有的 tRNA 分子都在末端腺嘌呤上附着氨基酸（图 114）。在分子的 5′末端总是发现有鸟嘌呤残基，但其功能不明。总之，一切 tRNA 分子都能够识别 mRNA 分子的核苷酸顺序这是毫无疑义的。在下面讨论遗传密码时，这一含义的重要性显得更为突出。现在的看法是 mRNA 核苷酸上每三个碱基为一组，称为密码子。每个密码子专门对应一个氨基酸。tRNA 中识别密码子的部位称为反密码子。反密码子由三个能和 mRNA 上密码子形成氢键的互补核苷酸组成。tRNA 分子除了这些共同特性之外，每个分子还有专供氨酰-tRNA 合成酶识别的位置。目前尚不能明确指出 tRNA 分子的哪个部位是合成酶的识别位置，但既然每一种 tRNA 分子的结构是独一无二的，可以想象，找出这样的识别位置当无多大问题。

究竟是氨基酸具有识别 mRNA-核糖体复合物所必需的特性呢？还是 tRNA分子或者它们的复合物具有此种性能？这个问题最初是用下面的方法澄清的。某些氨基酸分子在附着于 tRNA分子时，能够改变化学结构，转变为别的氨基酸而不改变原来 tRNA 分子的性能。例如半胱氨酸经阮来镍 (Raney nickel) 处理会变成丙氨酸(图 118)，从而亦使半胱氨酸-半胱氨酸 tRNA 复合物 (Cys-$tRNA^{Cys}$) 改变为丙氨酸-半胱氨酸 tRNA 复合物 (Ala-$tRNA^{Cys}$)。最后发现，这个变来的丙氨酸居然被插到了原先应为半胱氨酸在肽链的位置上。这个事实意味着氨酰-tRNA 之识别 mRNA，必定是属于 tRNA 的特性，而不是它所携带的氨基酸。

这一过程中颇堪玩味的问题是，所有的 tRNA 分子都具有 CCA 3′末端，而专一性的氨基酸附着到 tRNA 分子上都是通过腺嘌呤残基的 3′羟基和该氨基酸之间非专一性的成酯反应进行的。那么，氨基酸-tRNA 的专一结合，必定只有要求氨酰-tRNA 合成酶能同时识别氨基酸和 tRNA 分子，而且每一个酶都有与之相对应的一个氨基酸和一种 tRNA。然后，合成酶再以一般的成

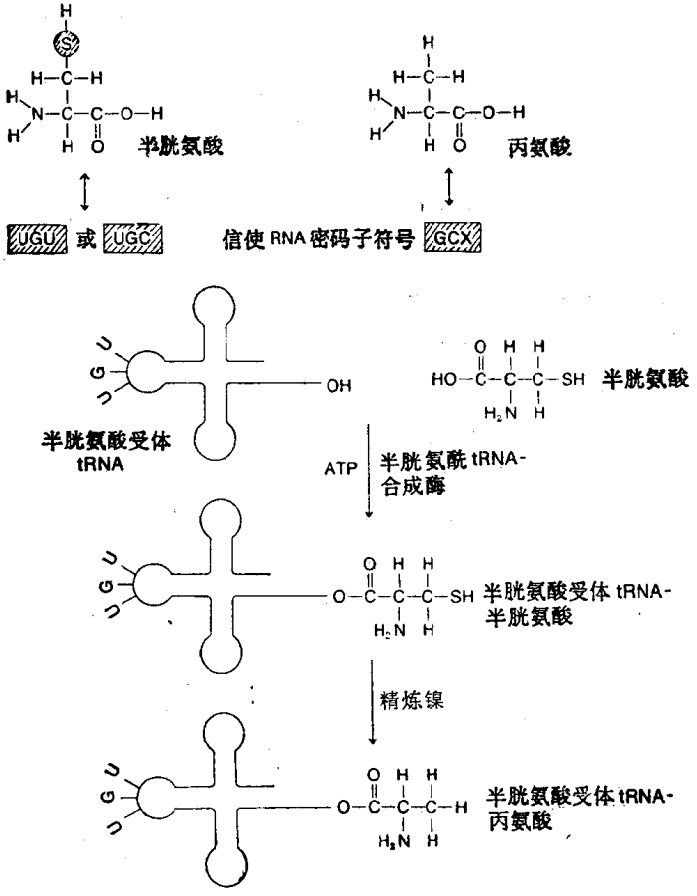


图 118 氨酰-tRNA 复合物的专一性

设计本实验的目的是用来判断 tRNA复合物所表现的专一性究竟是存在于复合物的氨基酸部分呢？还是 tRNA部分。测定专一性的方法是把氨基酸插入生长着的肽链，结果清楚表明，专门和半胱氨酸对应的tRNA，在连接丙氨酸时照样是把丙氨酸接到正常情况下半胱氨酸应占据的位置。因之，专一性系取决于tRNA分子上的反密码子部分。

酯反应使氨基酸附着于 tRNA 分子上。

tRNA 分子的结构和功能与其他的 RNA 分子不同，这一事实曾经引起过如下的争论：这些分子究竟是以 DNA 为模板合成的呢？还是以其他的化学方法形成的？由于最终分离出了含有一个或几个突变 tRNA 分子的突变型细菌，并且最近又证实细菌染

色体片断(基因)具有和大多数 tRNA 分子互补的碱基顺序，这才使问题稍有眉目。但是，即便作为"tRNA 基因"的证据已很明显，关于 tRNA 分子如何合成的问题的争论还是没有停止。

分析缺失核糖核酸酶 P(RNase P)的突变型噬菌体为澄清这个问题提供了主要线索(T_7 和 T_4是大肠杆菌的噬菌体)。核糖核酸酶是一类切割 RNA 链的酶，用核糖核酸酶突变型噬菌体感染大肠杆菌，寄主细胞不产生正常噬菌体特有的 tRNA 分子，而代之以分子量较大的 RNA 分子。事实上，这些 RNA 的分子量要比正常 tRNA 分子的分子量二倍还多，检测这些分子的结构（图119)，原来它们是小分子的二聚物。这个发现暗示着核糖核酸酶P的作用是在特殊的部位把二聚物切成二个有功能的 tRNA 分子和一个小的片断，由之获得的进一步启发是这二个 tRNA 基因在染色体上的位置应该彼此相邻，并且作为一个产生 RNA 的转录单位而行使功能。双基因转录本合成以后，再由核糖核酸酶加工成 tRNA。

tRNA 二聚物前体分子在被核糖核酸酶P切开之前，其中的二个 tRNA 分子虽尚未获得功能活性，但实际上已包括有几近全数的已知稀有碱基。由于在现有的 RNA 分子中只有这种二聚物分子以及有功能的 tRNA 含有这类稀有碱基，所以这些碱基看来不象是在合成时结合到二聚物上去的。现在已经证实，在合成以后和切开以前，在酶促作用下，普通碱基即直接转变为稀有碱基。

总的说来，tRNA 分子是以一切 RNA 分子固有的机制合成

图 119 tRNA 转录以后的加工

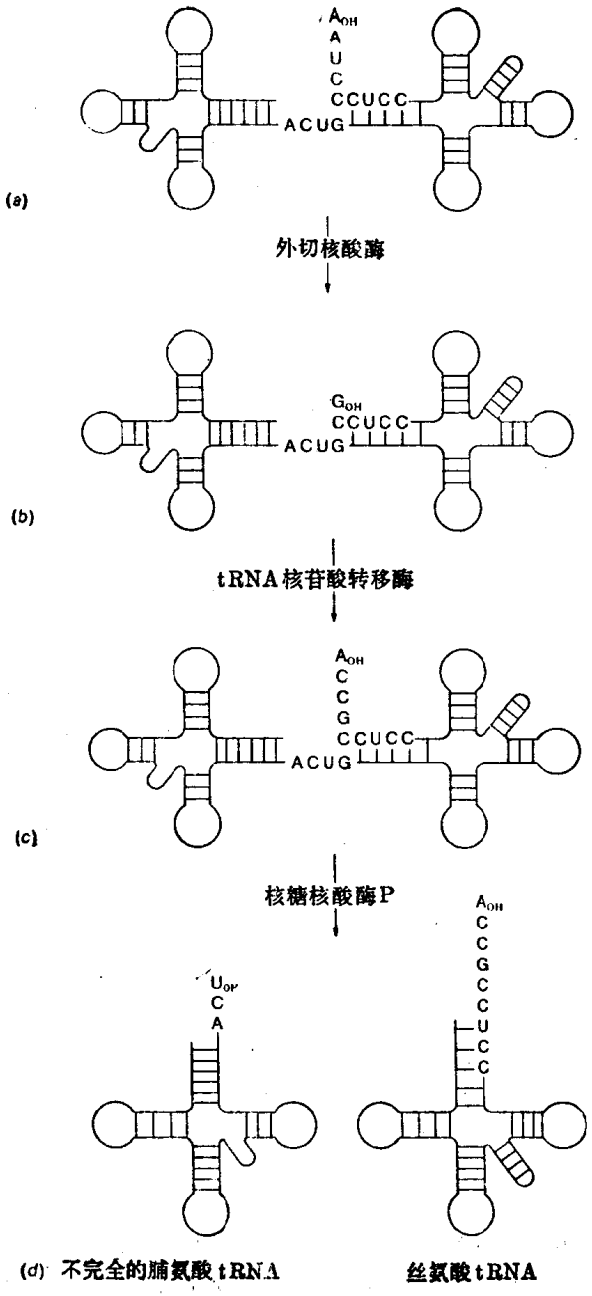
关于T_4噬菌体较大的前体RNA分子加工成 $tRNA^{Ser}$和$tRNA^{Pro}$ 的情况。

(a) 一个较长的RNA分子折迭以后，显示出是$tRNA^{Pro}$(左)和$tRNA^{Ser}$的前体分子。此时两个部分都不具有CCA末端。

(b) 外切核酸酶切除前体分子丝氨酸部分3′末端的UAA。

(c) tRNA核苷酸转移酶把CCA末端加到同一个 3′末端。

(d) 核糖核酸酶P在二个分子之间把链切开（切在脯氨酸部分的 3′U 和丝氨酸部分的 5′G之间)，$tRNA^{Ser}$加工完成，准备参加激活反应。$tRNA^{Pro}$仍需加上 CCA3′ 末端。(摘自 J. G.Seidman 与 W. H. McClain, Proc Natl Acad Sci USA, 72:1491, 1975。)


(a)
A
A
U
C
CCUCC
ACUG
外切核酸酶
(b)
G
CCUCC
ACUG
tRNA核苷酸转移酶
(c)
A
C
C
G
CCUCC
ACUG
核糖核酸酶P
U
C
A
A
C
C
G
C
C
U
C
C
(d) 不完全的脯氨酸tRNA
丝氨酸tRNA

的。它们一经合成，在蛋白质中起作用之前，还得经历一个分子加工的程序：第一步是将许多原来的碱基转化成稀有碱基，然后再由核糖核酸酶P把二聚物切成二个有功能的 tRNA 分子。目前已能从被核糖核酸酶P突变型病毒感染的细胞中解离出二聚物分子，当用从野生型噬菌体感染的细胞中提取的正常核糖核酸酶P处理二聚物后，由之产生的有功能 tRNA 分子可以说与体内产生的完全相同。有的二聚物分子产生两个相同的 tRNA 分子，有的也会产生两个不同的 tRNA 分子。如 $tRNA^{Pro}$ 和 $tRNA^{Ser}$ 就是从同一个二聚物分割而成的。

核糖体 RNA

这类 RNA 为核糖体所固有，所以命名为核糖体 RNA (rRNA)(图 120)。核糖体有二大类，即原核生物核糖体和真核生物核糖体。确实，核糖体大小的差别能够作为区分原核生物细胞和真核生物细胞的标志之一。两大类核糖体在蛋白质合成中所起的作用是相同的，但是每类核糖体的 RNA 和蛋白质在大小和组成方面却变化很大。

细菌的核糖体是重量为 3×10^6道尔顿的颗粒，根据它们的沉降特性(图 120)，这些颗粒称做 70S 核糖体，在电子显微镜下测得这些核糖体的直径约为 180Å。将 70S 核糖体在低浓度的 Mg 离子溶液中离心，可分解成为 50S 和 30S 两种颗粒。每个 30S 颗粒均包含有一个 16S的 RNA 分子，它的蛋白质占自身重量的 35%。16S RNA 分子约有 1,500 个核苷酸。每个 50S 颗粒包含一个约由 3,000 个核苷酸组成的 23S RNA 分子和另一个 120 个核苷酸的 5S RNA 分子(表 14)。

在多种细菌、植物和动物中现已鉴别出某些特定的据以转录产生 rRNA 的 DNA。细菌中转录成 16S 和 23S RNA 的基因呈重复排列(每个基因约重复 40 次)，并且集中在一段染色体中。重复 DNA 将在第六章中进一步讨论。但是这儿先提一下，细菌中为 5S、16S和 23S RNA以及tRNA、核糖体蛋白质的编码基因所占的染色体部分约为整个基因组的百分之 0.14 至 0.5。其中对应于

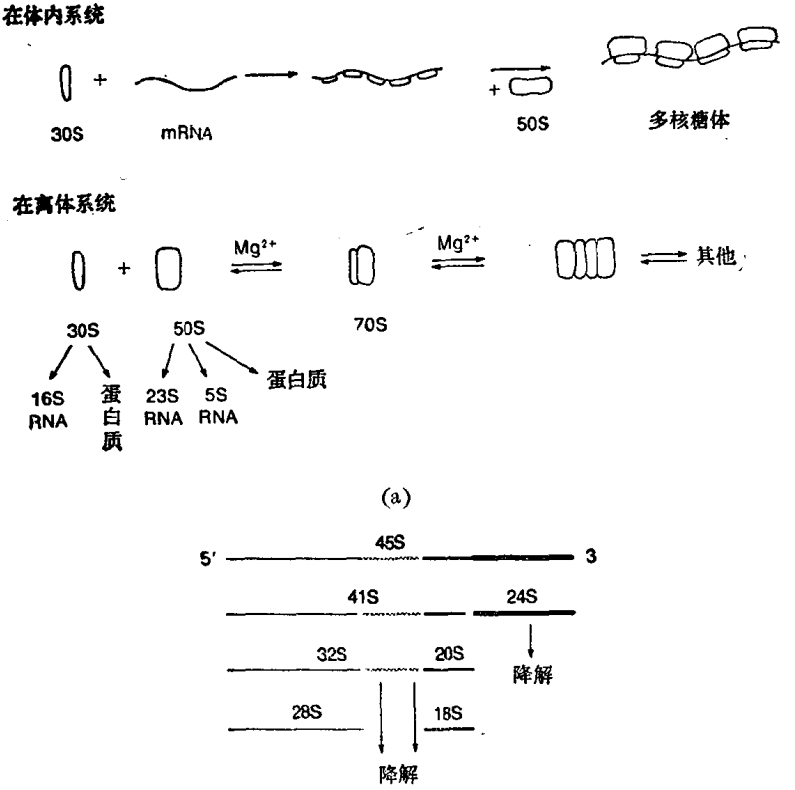


图 120 核糖体

(a) 从活的细菌细胞分离出来的核糖体经常聚集成多核糖体(上)，多核糖体是附着于同一个 mRNA 分子的一群核糖体。从多核糖体分离出来的单个核糖体是一个 70S的颗粒(真核生物核糖体是 80S的颗粒)。每一个 70S 的颗粒又能解离成一个 50S颗粒和一个 30S 颗粒(下)。 50S 颗粒由一个 23S RNA(rRNA)，一个 5S RNA以及共约 34 种蛋白质组成，30S颗粒由 16S RNA和 23 种蛋白质组成。30S颗粒能够完全解离成RNA和蛋白质单体，然后，离解的单体还可再重新聚合成有功能的核糖体。

(b) 哺乳动物(或许所有的高等动物)中的rRNA合成时是一个分子，它的沉降值大于小的核糖体亚基。图示 45S RNA降解产生 28S和 18S分子的步骤。

16S 和 23S rRNA 基因(rDNA)的重复竟高达 40 次之多，而且整段染色体是这样调节的：其中一个基因转录，所有的基因也都随之转录。虽然不了解其中的操纵机制，但它与蛋白质的合成速率密

表 14　关于核糖体结构、功能和大小以及某些大分子化合物大小一览表

细胞成分	结构	功能
核糖体	球状结构，由40%蛋白质和60%RNA组成	和mRNA结合后，把tRNA运送来的氨基酸相互连接，蛋白质由此产生
多核糖体或多聚核糖体	由mRNA暂时聚集的一些核糖体形成的线状结构	蛋白质合成的实际场所

成分	沉降系数	分子量(道尔顿)
核糖核酸酶(牛)	1.87	13,683
烯醇酶(酵母菌)	5.9	67,000
焦磷酸酶(酵母菌)	4.4	63,000
葡糖磷酸变位酶(兔肌肉)	3.69	74,000
NADPH*氧化还原酶(酵母菌)	5.82	102,000
葡萄糖-6-磷酸脱氢酶(酵母菌)	9.3	200,000
谷氨酸脱氢酶(牛)	26.6	1,000,000
30S 亚基 RNA	16	500,000
50S 亚基 RNA	23	1,000,000
5S RNA	5	(120个核苷酸)
小的核糖体亚基(大肠杆菌)	30	850,000
大的核糖体亚基(大肠杆菌)	50	1,800,000
完整的核糖体(大肠杆菌)	70	2,800,000
tRNA	4	(约80个核苷酸)

* 烟酰胺腺嘌呤二核苷酸磷酸(还原型)。

切相关，蛋白质合成时，所有的核糖体基因转录成RNA，而细胞处于蛋白质合成静止期时，这些基因即停止转录。

有人曾提出如下的旨在描述50S核糖体生成的可能程序：最早看到的50S颗粒的前体是一个其中约有60%碱基已甲基化了的23S RNA和3个核糖体蛋白质分子组成的复合物，该复合物的

沉降系数为 30S。接着 30S 颗粒增加，另外 9 个核糖体蛋白质分子成为 40S 的颗粒。在 23S RNA 的碱基全部甲基化和再增加七个以上蛋白质分子及一个 5S RNA 以后，便成为一个 50S 颗粒。这个 50S 颗粒还没有包括已知和 50S 核糖体缔合的全部 34 个蛋白质分子。在模拟条件下装配 30S 核糖体目前已获成功，它含有一个 16S 的 RNA 分子和大约 23 个不同的蛋白质分子，有关内容参见第五章。

有几种核糖体蛋白质研究得相当详细，30S亚基的蛋白质命以从 S_1 到 S_{23}(S＝较小的亚基)的名称，而 50S 亚基的蛋白质的名称则从 L_1 定到 L_{34}(L＝较大的亚基)。S_1 蛋白质连接在 16S rRNA 的 3′末端，3′ 末端的最后 49 个核苷酸能和多核苷酸分子的其余部分相分离，故可对这个片段进行单独研究。该片段形成类似于发夹的二级结构。有人认为正是发夹部分的核苷酸序列包含了和 mRNA 分子互补的碱基，在蛋白质合成开始时，互补碱基相互配对。在把 S_1 蛋白质从 30S 核糖体去除的情况下，便不能生成初始的复合物。

现在能够把完整的核糖体分解成游离的 RNA 和蛋白质亚基，然后，再把它们重新装配成有功能的核糖体。在进行分解和重新装配这二个过程中，都存在着蛋白质拆卸和聚合的一定程序。或许是最初有一个或几个蛋白质附着到 RNA 分子上，改变了核酸-蛋白质复合物的三级结构形态，从而为以后的蛋白质分子附着提供识别位置，依次重复，直到所有的蛋白质分子聚合完毕。已经鉴别出几个有特殊功能的核糖体蛋白质。例如，抗生素的识别位置蛋白，有酶功能的蛋白以及作为蛋白质合成起始和终止必需因子的蛋白。至于 rRNA 除了作为核糖体蛋白质的骨架以外，还赋予其他的功能，目前尚属未知之列。

真核细胞 rRNA 的结构更加复杂，首先这些 RNA 分子较大(18S 和 28S)，18S 和 28S RNA 分别和蛋白质结合成 40S 和 60S 的核糖体颗粒，完整的真核细胞核糖体的沉降率是 80S。转录时，18S 和 28S RNA 分子是当作一个重约 45S 的 RNA 分子而生成

的。45S RNA 合成以后，切成 18S 和 28S 二部分以及一个以后消失的片段。在切割反应以后，18S 和 28S RNA 分子的许多碱基即完成甲基化。

真核生物产生 18S 和 28S RNA 的基因集中在位于一个或几个染色体上的核仁形成区。在某些动物中，一个染色体上核仁形成区的每个基因重复达 800 次之多(有时 rDNA 基因集中的DNA区域比整个细菌染色体还要长)。由于这个区域的浮力密度和染色体其余部分 DNA 的浮力密度不同，所以能从染色体上分离下来。例如在爪蛙(*Xenopus*)中，rDNA 的 GC 含量较高，使得这部分 DNA 的浮力密度高于其余的 DNA。通过分离 rDNA 区段，可以看到其中的一条单链能和 18S RNA 或 28S RNA 杂交，从而显示出 rDNA 与 rRNA 的互补性。然而核仁形成区只有 50% 的 DNA 和 rRNA 互补，其余 50% 的 DNA要末不转录 rRNA，要末转录成 rRNA 后，又在 rRNA 分子加工期间消失。可见染色体的每个核仁形成区是由重复 800 次的 rDNA 基因和数量相当而功能不详的 DNA 组成的。"无功能"DNA 含有 73% 的 GC 碱基对。该区域内，两种功能不同的 DNA 的实际顺序尚未搞清，目前正在分析之中。

真核生物染色体上的 rDNA 共计占整个 DNA 的 0.05～0.5%，较为高等的有机体其百分率亦较高。但酵母菌却是一个例外，虽是低等真核生物，却有 2.4% 的 DNA 构成为 rDNA，而且已知酵母菌细胞的 RNA 与 DNA 之比也特别高。但是，总的说来，低等有机体 rDNA 基因数约为 100 个，或者更少些。染色体较多的动物一般具有 1,000 个 rDNA 基因，而高等植物的 rDNA 基因数则高达 10,000 个左右。还有一种原始的哺乳动物肺鱼(lungfish)(肺鱼不是哺乳动物，此处不确——译注)，它的基因组相当于大多数哺乳动物的 20 倍，共包含 5,000 个 rDNA顺反子。

从拍摄到的正在转录的 rRNA 电子显微镜照片中可以看到核仁形成区的超微结构(图 121)。图中无活性 DNA 一段一段规则地分布在活性单元之间，每个活性单元都含有一连串表现出固

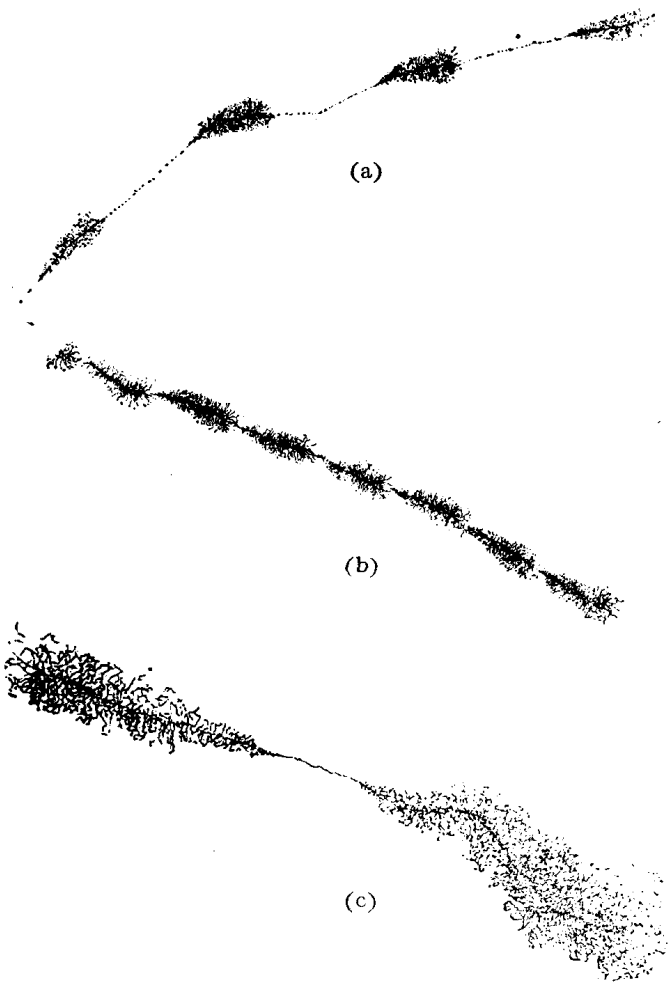


图 121　显示转录过程的电子显微镜图象

(a) 4个处于转录期间的rDNA基因电子显微镜图，显示出二个基因之间有很大的间隔。

(b) 从电子显微镜照片上可看到七个转录期的基因。基因之间没有明显的间隔。因为二张照片都是绿藻（*Bataphora oerstedii*）核仁形成区的电子显微镜图，所以间隔的作用不显著。

(c) 图示二个邻接基因，它们的转录方向相反。电镜照片系由整个胞核取样制成，有人推测这是结构基因正在转录为mRNA。（蒙德国 Dr. Sigrid Berger惠赠，Max-planck-lustitut, Wilhelmshaven，德国。）

有和均匀间隔的纤丝（多核苷酸链）。这些纤丝的长度各不相同，最长的纤丝表示将成熟的 RNA 分子。然而最长的 rRNA 分子还是短于它们据以转录的基因，有证据说明，它们是在从 DNA 模板释放之前，由于附着到蛋白质上而被截短的。在电子显微镜下所看到的所有 rDNA 活性单元上的纤丝，都表现为大体雷同的长度梯度，并且还显示出有一个多聚酶分子正附着于 rRNA 纤丝基部的 DNA 链上。使用更高倍电子显微镜，能够直接看到这一图景。

六十年代初，发现线粒体和叶绿体亦含有 DNA，这个发现引导人们对这些细胞器的遗传自主性重新作了研究，甚至有人推测不能排除它们起源于某些原始的原核生物的可能性。尤其是在发现叶绿体和线粒体等本身同样亦具备包括核糖体在内的蛋白质合成系统时，此说盛极一时。后来又发现一些引人入胜的事实，这些细胞器中的核糖体和细菌的核糖体不相上下，反而与真核生物细胞质中的核糖体大相径庭。细胞器核糖体的 rRNA 分子和细菌的 rRNA 分子的数量级相同，而且细胞器核糖体对已知能使细菌核糖体失活的抗生素同样是敏感的。有关细胞器核糖体和细菌核糖体之间在重量大小上的雷同性现概要列于表 15。

表 15　核糖体分类

	真核生物	原核生物	细 胞 器
核糖体	80 S	70 S	60～70 S
大亚基	60 S	50 S	～50 S
小亚基	40 S	30 S	～30 S
小亚基 RNA	18 S	16 S	～16 S
大亚基 RNA	28 S	23 S	～23 S
5S RNA	+	+	−

信使 RNA

在论述 mRNA 之前，先简单介绍一下有关说明蛋白质合成经

常使用的几个术语。作为 tRNA 和 rRNA 模板的 DNA 被称为 tRNA 和 rRNA 基因，这些基因和转录 mRNA 的基因(结构基因）的区别在于只有结构基因才能把信息传递给蛋白质。迄今为止尚未发现任何蛋白质的氨基酸顺序反映了 tRNA 和 rRNA 的核苷酸顺序，它们的功能主要是协助蛋白质的合成，而蛋白质中的氨基酸顺序则又是由依照结构基因转录成的 mRNA 控制的。

Jacob 和 Monod 首先深刻地揭示，包含在结构基因中的信息乃是以互补的方式转录成 RNA 碱基顺序，他们把这种 RNA 转录本称为信使 RNA(mRNA)。实际上，早在 Jacob 和 Monod 提出信使 RNA 理论之前，就有人分离成功这种今天在化学上定义为 mRNA 的 RNA，但当时没能正确理解它们的生物学含义。

mRNA 理论发表之后，人们即开始通力研究 mRNA 分子，所得普遍规律如下：(1) mRNA 分子的大小互不相同，正与肽链的长短多种多样相吻合。(2) 与 tRNA 和 rRNA 相比，mRNA 在细菌内的存留时间较短，周转非常迅速，一般每种类型 mRNA 只有几个分子同时存在。(3) mRNA 分子的碱基顺序只和结构基因二条链中的一条链互补，mRNA 与 tRNA 基因及 rRNA 基因无互补关系。

由于多肽长度各有不同，可以设想 mRNA 的长度也有参差。但是 mRNA 分子长度的变化还产生于某些 mRNA 分子是从一个以上的基因转录而成的（这些 mRNA 称为多基因或多顺反子信使)。例如大肠杆菌有一个转录 5 个基因的多顺反子 mRNA(这些基因编码 5 个有关色氨酸合成的酶)，其长度为 7320 个核苷酸，已知这个特定的 mRNA 分子的转录速率是每秒钟 50 个核苷酸(37℃)。在鼠伤寒沙门氏菌(*Salmonella typhimurium*)中，一个这样的 mRNA 分子是 10 个基因的产物。这 10 个基因编码 10 个和组氨酸生物合成有关的酶(这种合成方式以后将再介绍)。还分离到了其他的多顺反子 mRNA。但是迄今为止，只在原核生物中观察到有多顺反子 mRNA 的存在。

在正常生长的大肠杆菌细胞中，可以看到转录大约在 500 个

独立的位点同时进行，其中约200个位点对应于较稳定的 rRNA 分子和 tRNA 分子。它们拷贝的基因占基因组的0.7%左右，即相当于 1.6×10^5核苷酸。其余300个位点担负着mRNA 的合成，它们转录的基因将近基因总数的10%（即不超过 2×10^6 个核苷酸）。在任何时刻，基因组内的大多数DNA 没有转录活动。同样，无论何时，整个 RNA 能划分为97%稳定的 tRNA 和 rRNA 以及3%的 mRNA。而在 mRNA 中，约有70%其据以拷贝的基因仅占整个基因组的1%（2×10^5个核苷酸对）。有关 mRNA 的细节将在遗传密码探讨之后进一步阐述。

遗 传 密 码

知道遗传信息从 DNA 转录成 RNA 和信息再由 RNA 转移到蛋白质是一回事，而破译遗传信息则是完全不同的另一回事。Watson和 Crick 发表DNA 结构模型之后，分子遗传学家们便一致同意遗传信息由4个字母 A、T、G 和C 组成。后来 Jacob 和 Monod 又进而提出了信息从结构基因(DNA)转录到 RNA 的概念，大家赞成 RNA 信息的字母表除了T要由U来代替之外，也应和 DNA 相同。人们早就知道天然存在的氨基酸有20种，但是直到搞清楚 DNA 的分子结构之前，大多数遗传学家没有能以4个字母的20种组合来解释这个事实。1953年～1960年间，在许多论文中已试图找出把4个字母排成20个字母组合的种种方法，但直到1960年，这个问题才由实验加以解决。

如果从4字母表中仅取二个字母组成一个“词汇”，能产生16(4^2)个词汇，如果取三个字母组成一个“词汇”能产生64(4^3)个词汇，这里，一个词汇也就是一个信息单位(图122)。假定每个氨基酸由一个 DNA 词汇表达，那么照第一种排法，总数为16个信息单位，单位数比起码需要的20个氨基酸词汇要少4个。但是第二种排法则有64个信息单位，即便有几个作为“标点”的话，总数却是20个氨基酸的3倍，又显得太多。六十年代初期，二个大不相同的实验，几乎不约而同得出了三个核苷酸编码一个氨基酸的结

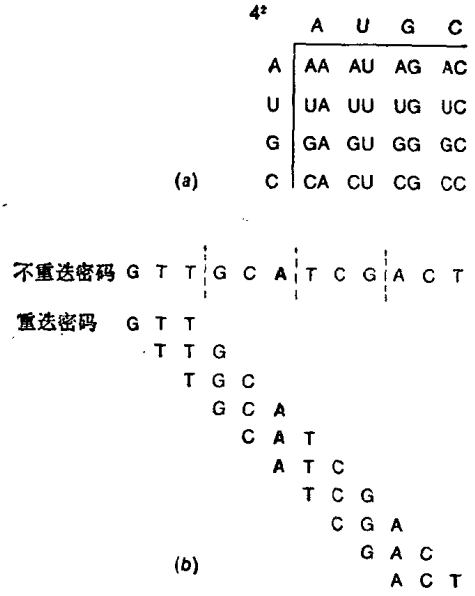


图 122 遗传学词汇编码法

(a) 4个字母(A、U、C、G)的字母表以2个字母编码一个词汇，将产生16(4^2)个词汇，同一个字母表用三个字母编码，则产生64(4^3)个词汇。

(b) 若密码互相重迭，每个字母就能参与组成三个密码词汇，信息单位的总数量将大为增加。

论，这种三位一体的核苷酸编码称做密码子。

Crick, Sidney Brenner 和他们小组做的一个实验是利用遗传学的手段来探索编码方法。早先 Benzer 曾培育出一个 T_4 噬菌体缺失突变型，这个突变型在 r_{II} 区域缺失的片段包括顺反子 A 和 B。然而，正如早已指出的那样，即使顺反子 B 部分缺失，它的功能依然存在。但顺反子 A 存留部分的突变却会使顺反子 B 失去功能(见图 93)。如果认为转录只是从左至右认读密码，那么碱基对的增加或缺失将会改变从突变点起直到顺反子结尾止的读码节段(reading frame)。Crick 首创的术语移码突变 (frame-shift) (+)和(−)便是改变从突变点起顺次而下的读码节段的每个密码子的成分(如图 123 所示)。若发生移码突变，原有读码节段的信息就被更替，仅当读码节段回复到原来位置(节段)，它的原义才能恢

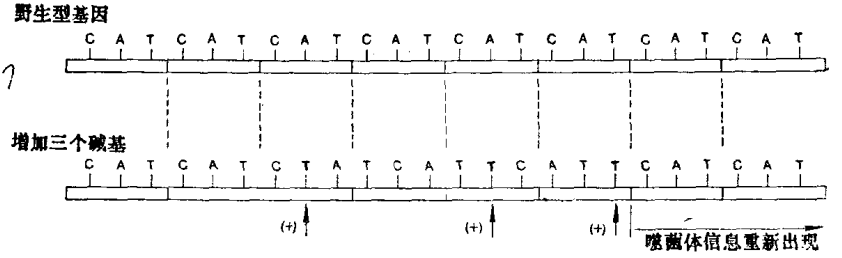


图 123　读码节段与移码突变

吖啶类染料诱发DNA的碱基对增加或缺失（图92）。如果基因的全部碱基从一端到另一端被认读，那么碱基的增失都将改变原先的读码节段，也就是(+)突变的读码节段可为(−)突变回复。反过来也是一样。
欲知(+)突变是否能为增加碱基所回复？如能回复，要增加几个核苷酸？同样(−)突变的回复条件如何？可这样来检测：如果密码子的长度为三个核苷酸，可以期待如加上三个或缺失三个核苷酸都有可能回复原先的读码节段。

复(图 123)。

如果造成的无意义产物的单个(+)突变由一个(−)突变来校正，这显然意味着节段经过一段摇摆后复位，并在余下的整个区段中继续发挥它们原先的作用。为使读码节段能够回复到它的原先状态，就必须要丢失或增加的核苷酸总数为密码词汇的倍数，这时(+)突变型可由增加的(+)突变来矫正，(−)突变型同样由(−)突变来矫正，根据(+)或(−)突变的数目就可定出密码词汇的长短。例如，密码子的长度是二个核苷酸，凡是 2 个、4 个或 6 个等等双数的(+)突变都可能回复原先的读码节段。如果密码子的长度是三个核苷酸，那么凡是 3、6、9 等数目为 3 的倍数的(+)突变也应该会使原先的读码节段复位。

吖啶类染料能够诱导大量的(+)和(−)突变，利用遗传学方法能够获得许多(+)和(−)数目不同的突变型菌株。结果发现只有 3 的倍数的(+)或(−)突变型才能恢复读码节段的原义，而任何其他数目的(+)或(−)突变体只产生无意义的产物。换言之，这些结果也同样有力地支持了三联体密码，即每三个核苷酸为一个密码子的结论。

探索编码问题的第二种方法，是 Nirenberg 和 J. Heinrich

Matthaei 使用的生物化学方法。他们在体外蛋白质合成系统中以人工合成的 mRNA 作为模板，然后测定其中产生的多肽的氨基酸顺序。这种方法产生的人工多肽能和天然 mRNA 分子产生的多肽相比较。前面记述过 Ochoa 利用多聚核苷酸磷酸化酶合成的人工mRNA 并没有固定的顺序。但这个无专一性的缺陷，能以控制磷酸化酶的底物，即核苷酸浓度的简单方法得到一定程度的弥补。例如，以 4:1 的比率加入A和U，得到各种三联体的机率如下：

$$\text{AAA}\ 64=\frac{64}{125}$$

$$\left(2\text{个A}+1\text{个U的机率为}\frac{16}{125},\text{共三种}\right)$$

$$\text{AAU、AUA、UAA}\ 16\times3=\frac{48}{125}$$

$$\left(1\text{个A}+2\text{个U的机率为}\frac{4}{125},\text{共三种}\right)$$

$$\text{AUU、UAU、UUA}\ 4\times3=\frac{12}{125}$$

$$\text{UUU}\ 1=\frac{1}{125}$$

以这种混合底物合成的mRNA分子为模板，可以期待转译的多肽链的氨基酸频率将和模板三联体密码的频率相符合。

Nirenberg 和 Matthaei 的早期实验包括使用多聚尿嘧啶核苷酸(多聚U)作为探索工具。结果证明多聚U在所使用的系统中是有功能的，它转译的多肽是多聚苯丙氨酸(图 124)。同样，多聚A mRNA 产生多聚赖氨酸，多聚C 产生多聚脯氨酸，多聚G产生多聚甘氨酸。UUU, AAA, CCC 和GGG 是从 4 个字母排成的 64 种三联体密码中的四种，以这样的办法能够毫不含糊地测定这些密码的功能。

使用无专门顺序 mRNA 分子作出的统计学结论包含机率问题，为了确切无误地探索这一问题，Nirenberg 和 Philip Leder

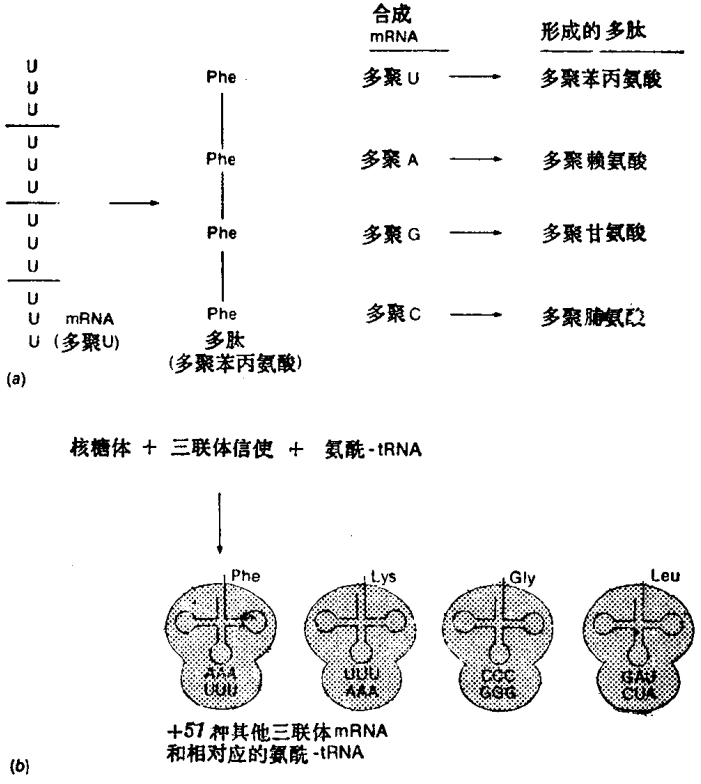


图 124 合成 mRNA 和遗传密码

(a) 由一种核苷酸合成的mRNA，例如多聚U，在体外蛋白质合成系统中转译成由一种氨基酸聚合成的多肽，例如多聚苯丙氨酸。这些实验可告诉我们某些密码子的核苷酸组成，但不能说明长度。

(b) 长度为三个核苷酸的mRNA分子，能揭示密码子的组成和长度。Nirenberg和 Leder 用这种方法鉴测了各种密码子，并发现一种或几种氨酰-tRNA 复合物能够识别 61 种密码子中的一种密码子，另外还有三种密码子不能被任何的氨酰-tRNA识别（现在知道这些密码子正是转译的终止信号）。

又发展了另一种方法。他们合成了全部可能排成的 64 种密码子，即三个字母的“mRNA”分子，然后把三个字母的 mRNA 和核糖体构成 mRNA-核糖体复合体。在复合体中，加进“活化”氨基酸：氨酰-tRNA 分子。由于带有特定氨基酸的氨酰-tRNA 分子与人工编码的（已知的）mRNA 核糖体的专一性结合，从而最终揭示了

64 种密码子中的 61 种密码子和 20 种氨基酸之间的对应关系（图 124），这里除了澄清了密码子-氨基酸的指配关系 (assignment) 外，还观察到有关遗传密码的其他一些特性：

1．遗传密码是简并的，即在某些场合几种密码子识别相同的氨基酸。

2．所有的密码子都是三联体。

3．64 种三联体密码中只有 61 种能够编码氨基酸（其余的三种密码子从未发现与任何活化氨基酸有对应关系，现在知道这三种密码子乃是转译的终止信号，即等于是标点）。

4．一种密码子在一切有机体内编码同一个氨基酸，换言之，遗传密码是普遍通用的（图 125）。

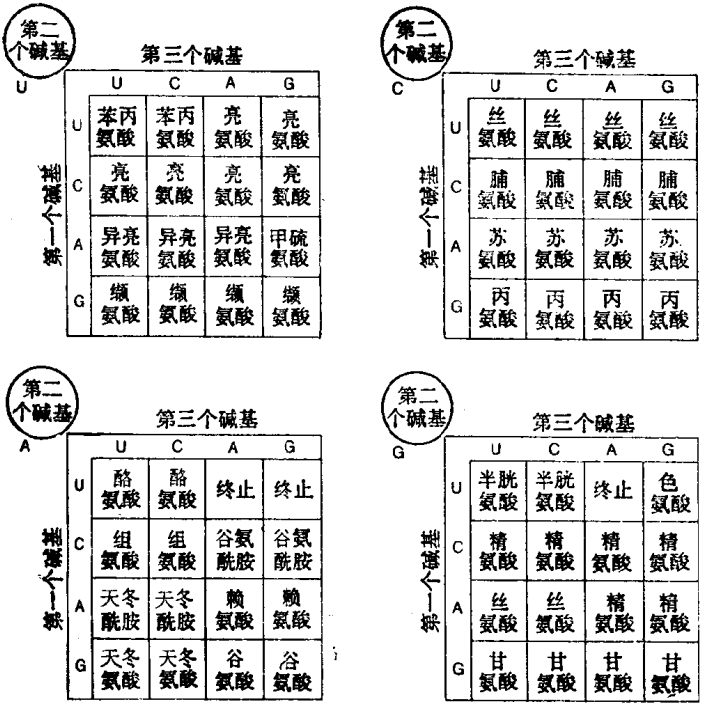


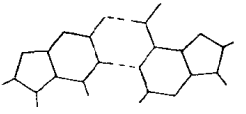
图 125　遗传密码

图示 64 种三联体密码子和它们编码的氨基酸。UAA，UAG和UGA不编码任何氨基酸。色氨酸仅仅由一个包含U和二个嘌呤的密码子(UGG)编码，三个不编码氨基酸的密码子已知是终止密码子。

照平均计算，每个氨基酸应当有三个不同的密码子。但密码子的分配是不规则的，一个氨基酸对应的密码子从一个到六个不等。如果仔细观察三联体-氨基酸的指配关系(图 125)，将会看到在大多数情况下，简并密码是一些第三个字母不同的密码子。例如，有一类情况，在第三个位置上不论哪种嘧啶碱基都能编码同一个氨基酸。这一事实曾使 Crick 据以提出摆动假说 (wobble hypothesis，图 126)，强调三联体开头二个字母的顺序是严格的，第三个字母则允许变动。第三个位置可安放几种碱基，或者说可以摆动。摆动碱基的配对规则见图 126。

密码子第三个碱基的配对规则

反密码子碱基	识别的密码子碱基
U＋修改U	A G
C	G
A	U
G＋修改G	U C
I(次黄嘌呤见图47)	U C A



次黄嘌呤核苷　　腺嘌呤

图 126　摆动假说

多重密码 (code redundancy)现象(即几个密码子翻译同一个氨基酸)首先由Crick以摆动假说加以解释。这假说的命名是因为观察到对于密码子的专一性来说，其开头二个字母要求比第三个字母更加严格。也就是说，第 3 个位置可以摆动，而不会影响密码子的专一性。摆动规则如图所示。

遗传密码的简并向人们提出了有趣的课题，即如何从 tRNA 分子和氨酰-tRNA 合成酶的角度来认识简并的机理。一般认为，三联体核苷酸 mRNA 实验暗示出 mRNA 密码子是以简单而直接的互补反应识别 tRNA 反密码子的，其中包括在密码子和反密码子之间形成氢键。一旦能把各种 tRNA 分子的碱基顺序确定下

来，则其多种多样的反密码子必定会在 mRNA 上找到与它们对应的密码子并两相对号(表 16)。可见，如果在密码子和反密码子之间确实存在简单的互补关系，那末密码子的简并必须反映在反密码子中，也就是反映在 tRNA 分子中。这一设想促使人们去努力寻找那些能运载同一个氨基酸 tRNA分子间的不同。结果在大肠杆菌中就发现了 5 个亮氨酸受体 tRNA，3 个酪氨酸受体 tRNA 和 2 个丝氨酸受体 tRNA。每一种氨基酸的不同受体 tRNA 在其初级结构(核苷酸顺序)以及和氨基酸的亲和力上都各有千秋。

很显然，识别同一个氨基酸的 tRNA 与氨基酸亲和力的差别，可由存在着相应各别的氨酰-tRNA 合成酶来说明。至少已分离出二种识别同一个氨基酸的氨酰-tRNA 合成酶。这里存在几种可能情况：一是 61 种密码子各有其彼此不同的 tRNA，而每个 tRNA 同样各有其不同的氨酰-tRNA 合成酶；一是一种氨酰-tRNA 合成酶能够以不同的亲和力识别几种 tRNA 分子(表 17 表示 tRNA 合成酶的特异性，它能催化不同种属来源的 tRNA 分子，使之运载有关氨基酸，从而显示出功能通用性)。

表 16 mRNA 密码子和相应的 tRNA 反密码子*

氨基酸	密码子	tRNA 反密码子
丙氨酸	GCX	IGC
丝氨酸	UCX	IGA
缬氨酸	GUX	IAC
酪氨酸	UAP_Y	GψA
苯丙氨酸	UUP_Y	G'AA

* X＝U、C或A，　P_Y＝嘧啶，　G'＝2'-O-甲基鸟嘌呤，　ψ＝假尿嘧啶，I＝次黄嘌呤(一种嘌呤前体)。

某些氨基酸能被一个以上的 tRNA分子运载到生长着的肽链的结合位置上，这种现象称为密码冗余。表示其数量概念的用语叫多重数级或多重程度(degree or extent of redundancy)，两词

表 17　tRNA 合成酶的专一性

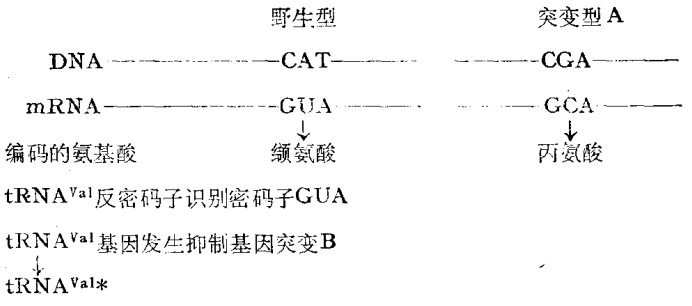
运载有关氨基酸的tRNA来源	有关氨基酸的tRNA酶的来源	tRNA 运载有关氨基酸的能力					
		甘氨酸	亮氨酸	赖氨酸	苯丙氨酸	脯氨酸	丝氨酸
大肠杆菌	大肠杆菌	H	H	H	H	H	H
	产气气杆菌	H	H	H	H	H	H
	鼠伤寒沙门氏菌	H	H	M	H	H	H
	Micrococcus lysodeikticus	H	H	H	H	H	H
	酵母菌	VL	L	H	L	0	M
	兔	VL	L	L	0	0	H
酵母菌	大肠杆菌	L	VL	L	L	L	0
	产气气杆菌	L	VL	L	L	VL	0
	鼠伤寒沙门氏菌	L	VL	L	L	L	0
	M. lysodeikticus	H	VL	L	M	H	L
	酵母菌	H	H	H	H	H	H
	兔	M	M	M	H	M	H
兔	大肠杆菌	0	0	L	VL	0	0
	产气气杆菌	0	0	L	VL	0	0
	鼠伤寒沙门氏菌	0	0	L	VL	0	0
	M. lysodeikticus	L	0	M	L	H	L
	酵母菌	H	H	H	M	H	H
	兔	H	H	H	H	H	H

* H(高)＝达到同源tRNA与tRNA合成酶结合作用时运载能力的 70%。
M(中)＝达到同源tRNA与tRNA合成酶结合作用时运载能力的30～70%。
L (低)＝达到同源tRNA与tRNA合成酶结合作用时运载能力的 5～30%。
VL(极低)＝达到同源tRNA与tRNA合成酶结合作用时运载能力的 5%。
0＝异种结合时无运载能力。

同义。例如，亮氨酸有五个受体 tRNA 就是一种密码冗余现象。另外在抑止基因突变型中也显示有密码冗余的存在。如果结构基因发生一个原突变，细胞或有机体的表型一般至少有二种"校正"方法。第一种是较为直接的回复突变方法，即第二个突变恢复了核苷酸原来的顺序。第二种方法因为历史上的原因称为抑制基因突变，有几种类型。抑制基因突变意即发生在 DNA 染色体不同位点上的第二个突变，使表型从突变型又重新回复到野生型。例如，T_4 噬菌体 r_{II} 区域弥补(−)突变的(+)突变，就是抑制基因突变。在这种情况下，(+)突变体以回复正确的读码节段的方式"抑制"了(−)突变，因此恢复了野生型表型。

另一类以部分或全部恢复野生型表型为特征的抑制基因突变，其抑制效果是通过改变专一的tRNA分子或相应的氨酰-tRNA合成酶的遗传特征而取得的（图 127）。如果引起一个氨基酸替代另一个氨基酸的染色体原突变合成一个无功能蛋白质，那么一个突变的 tRNA 分子也有可能把一个"突变"氨基酸插入原先将被无效氨基酸替代的位点。如果突变tRNA插入的氨基酸刚好是被原突变替代掉的氨基酸，那么即使原突变造成的突变型 mRNA 继续产生，也将合成一些野生型蛋白质。抑制基因突变对野生型表型的恢复程度有赖于抑制基因突变可能产生的野生型蛋白质的数量和细胞为获得野生型表型所需要的该蛋白质的数量。然而，这样的细胞无论如何也不会和野生型细胞完全相合，因为除了野生型蛋白质之外，它们还产生突变型蛋白质。这种突变型 tRNA 分子（或突变型合成酶）的突变校正能力可由野生型蛋白质与突变型蛋白质的比率来测定（图 127）。

现在已经知道 DNA、RNA 和蛋白质语言之间的真实关系以及核酸字母对蛋白质文字的密码翻译表。揭示这些编码关系的理论和实验构成了今日分子遗传学发展的基本内容。但是了解核苷酸和氨基酸之间的编码关系，并没有告诉我们转译是怎样起始、终止和调控的。确实，蛋白质合成的起始和终止问题已经超出了简单的编码关系，这里面包含有很多机理，如 mRNA 和核糖体形成



突变B产生的 $tRNA^{Val*}$ 有时以很低的百分率识别GCA，结果缬氨酸插入正常情况下与密码子GCA对应的丙氨酸的位置，这种包含缬氨酸的蛋白质将产生野生型或接近野生型的表型。

另一种影响氨基酸活化酶专一性的突变，同样能抑制突变A的作用。这种突变使携带丙氨酸的 $tRNA^{Ala}$ 偶而携带缬氨酸，从而产生与 $tRNA^{Val}$ 发生变化相同的结果。

图 127　突变型tRNA抑制原突变

遗传学上使用抑制基因突变一词，既沿用历史习惯，表达亦很方便。抑制基因突变用来描述 DNA 上与原突变位点不同的位点上发生的并且把原突变产生的突变表型回复到野生型的任何第二次突变。大多数情况，带有原突变的细胞或有机体由于同一个基因组内存在抑制突变，将表现为野生型。另外，也有抑制突变通过tRNA起作用的情况。例如，原突变导致丙氨酸取代肽链(酶)上的缬氨酸。这样的突变可为偶然发生的使 $tRNA^{Ala}$ 变为携带缬氨酸的 tRNA的突变所抑制。如果突变产生足够的具有缬氨酸的蛋白质，即使突变型蛋白质照样产生，细胞或有机体将表现为野［型。

复合物，mRNA-核糖体复合物识别氨酰-tRNA 分子，肽链的位置，附着和延伸，以及这些复合物的拆除和再循环等等。

蛋白质合成器

一切细胞在蛋白质合成开始之前，必须准备好蛋白质合成器(apparatus)。在这个机构的构成物中包括一个 mRNA-核糖体的复合物。这个复合物是由 mRNA 分子附着到核糖体的 30S 亚基形成的。为了完成附着，必须有一些称为起译因子的核糖体联结蛋白存在。起译因子和核糖体联系松散，它们在核糖体离解过

程中和在离体合成系统中制备时容易丧失。没有起译因子，核糖体不能和mRNA结合。细菌中有三种起译因子，它们在某种程度上乃是以物理特性而不是以功能为其标志特征的(图128)。

细菌中，蛋白质合成的起始过程相当复杂，但是仍能概括为一些简单明了的步骤。第一个和mRNA-30S核糖体连接的是甲酰甲硫氨酸-tRNA($tRNA^{fMet}$)(图128)。但在tRNA附着到mRNA分子之前，mRNA本身必须先附着到30S核糖体上。这个附着是合成起始的第一个步骤，包括mRNA的AUG密码子，30S核糖体和作用机制不明的起译因子IF-3等构成物。整个复合物称做蛋白质合成复合物I，用IF-3-30S核糖体-mRNA来表示。然后，$tRNA^{fMet}$就附着在这个复合物上。这时又伴随着另一个起译因子IF-2和GTP的结合产生复合物II：IF-2-30S核糖体-mRNA-$tRNA^{fMet}$-GTP。

构成蛋白质合成器的下一个步骤是50S核糖体附着到复合物II上。这个附着由GTP变成GDP相伴随，从中产生复合物III：70S核糖体-mRNA-$tRNA^{fMet}$。$tRNA^{fMet}$结合到这个复合物50S核糖体亚基的肽位(P)上。这一系列反应组成蛋白质合成的起始阶段。

复合结构III就是蛋白质合成器，它一旦形成，肽链的延伸也就随即开始。AUG是mRNA分子的起始密码子，与它相邻的第二个密码子是与$tRNA^{Ala}$对应的密码子。第二个氨酰-tRNA附着在复合物III 50S核糖体的A位。这个附着在延伸因子EF-T和GTP存在的条件下才能发生。

当A位和P位同时有氨酰-tRNA分子占据时，一个和50S核糖体相联系的肽基转移酶把P位的氨基酸和A位的氨基酸相连接，这个反应发生以后，"缺载"tRNA即从P位释放，mRNA分子连同A位的氨酰-tRNA向前移动，于是tRNA-二肽进入P位。一个新的氨酰-tRNA占据A位，而肽基转移酶再使二肽和第三个氨基酸结合。这个过程反复进行，直到合成终止。

蛋白质合成以三联体UAA、UGA或UAG(终止三联体密码)

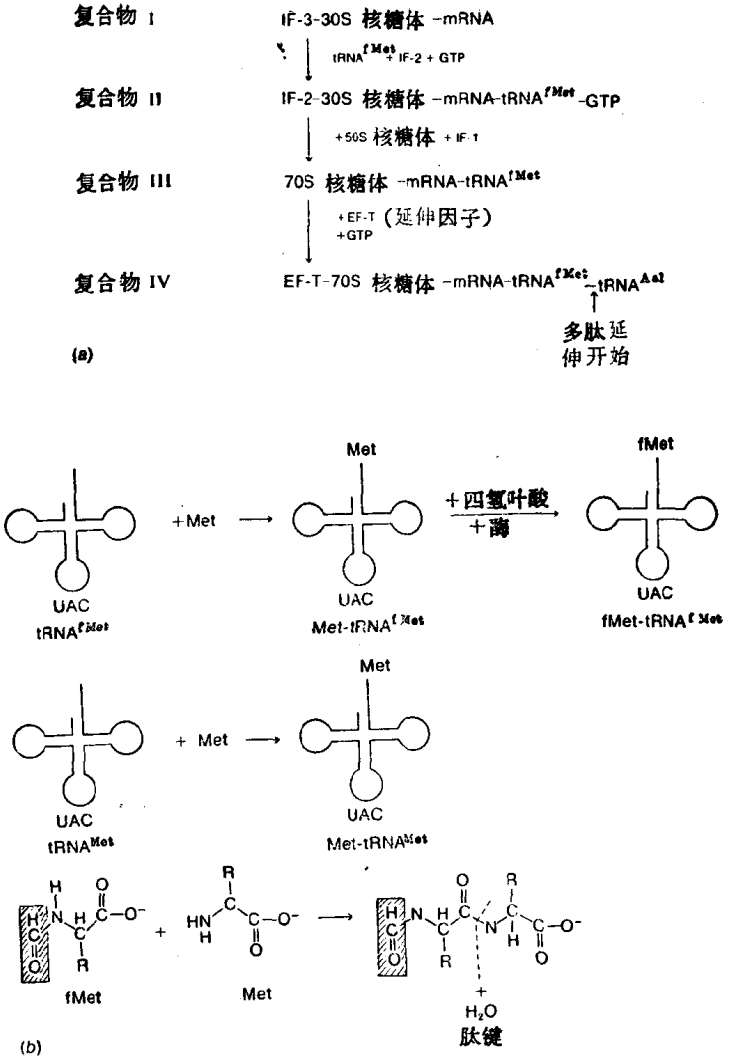


图 128 蛋白质合成器的装配

(a) 在细菌蛋白质合成器中，现在至少发现四种复合物，分离到三种与合成有关的起译因子。IF-3(起译因子3)是复合物 I 的组成部分，IF-2是复合物 II 的组成部分，IF-1以某种方式参与复合物III的形成。延伸因子(EF-T)和复合物IV相联系。

(b) 图示$tRNA^{fMet}$和$tRNA^{Met}$的运载反应。肽链N末端氨基酸是N-甲酰甲硫氨酸。图中第二个氨基酸是甲硫氨酸。甲酰甲硫氨酸的氨基为甲酰基所封闭，以致甲酰甲硫氨酸不能出现在肽链的内部。

中的一个或一个以上作为终止信号。另外至少还有二个终止因子，R_1 和 R_2。R_1 识别 UAA 和 UAG，R_2 识别 UAA 和 UGA，对于这些因子，目前还一无所知。

现在开始讨论另一个与起始有关的问题。N-甲酰甲硫氨酸附着的 tRNA分子和甲硫氨酸附着的 tRNA不同，虽然 $tRNA^{fMet}$ 和 $tRNA^{Met}$ 识别同一个密码子 AUG。当密码子 AUG 位于 mRNA 分子的内部，它吸引 $tRNA^{Met}$，而当 AUG 出现在 mRNA 的首端，则吸引 $tRNA^{fMet}$。差别的原因还不清楚，但是据信起始密码子 AUG 受到起始因子或者相邻的在细菌或病毒中几乎总是丙氨酸密码子的影响是一个原因。事实上，来自大肠杆菌和一些 RNA噬菌体的所有蛋白质，它们的氨基端，或者说起始端的氨基酸顺序是相同的，都是甲酰甲硫氨酸-丙氨酸-丝氨酸。有一些场合，N-端残基在肽链合成之后才解离，而另一些场合，fMet 和 Ala 在合成开始之后立刻被切除。在大肠杆菌中，大约 45％的蛋白质的N-端氨基酸是甲酰甲硫氨酸，30％是丙氨酸，15％是丝氨酸。然而，还不能肯定 AUG 相邻的丙氨酸和丝氨酸密码子是否对端部 AUG 识别 $tRNA^{fMet}$ 而不识别 $tRNA^{Met}$ 真有作用。

只有 mRNA-核糖体复合物的第一个密码子位置能容纳 $tRNA^{fMet}$。甲酰基使得甲酰甲硫氨酸对于除了第一个密码子之外的任何密码子都显得太大了，因而，合成一旦开始，$tRNA^{fMet}$ 便不能和复合物结合。这个事实使人们推测多顺反子 mRNA 分子在一个顺反子的终止密码子和下一个顺反子的起始密码子之间必须有一定的间隔，否则 $tRNA^{fMet}$ 无从使第二条肽链开始合成。

大肠杆菌有二种甲硫氨酸-tRNA，但只有一种能使甲硫氨酸甲酰化，即由甲酰基CHO 替代氨基中的H 原子。甲酰化是在甲硫氨酸附着到 tRNA分子之后再发生的。二种 tRNA 都识别 AUG，但 $tRNA^{fMet}$ 除此之外，还识别 GUG，因此 AUG 和 GUG 都能作为起始密码子。但在不作为起始密码子时只有 AUG 识别甲硫氨酸-tRNA，GUG 不处于起始位点时，则把缬氨酸结合进肽链。

概括地说，我们已了解到链的起始过程本身能为确定 AUG 是

否为起始密码子提供了判断方法，而且正是 AUG 才确保核糖体在mRNA上附着位置的正确无误。这个过程还说明了 30S 和 50S 核糖体亚基配置成 70S 核糖体的方式，及 70S 核糖体把肽基-tRNA分子放到核糖体上有利于 tRNA 反密码子识别 mRNA 密码子的适当位置上。

Howard Dintzis 利用简略地表示在图 129 的实验方法，证明蛋白质合成系从肽链的N端到C端的方向开展。他的办法是把 ^{14}C 标记的氨基酸加到血红蛋白肽链的生长系统中，隔一定时间直接观察标记氨基酸的结合情况。因为系统中时时都有新链开始合成和旧链合成终止，在任意时刻，中断合成过程结果发现放射性只存在于未完成链的N末端和成熟链的C末端。图 129 明确地表示了肽链的延伸方向是由N端到C端进行的。

在细胞的蛋白质合成旺盛时期，大多数 mRNA 分子都附着有好些核糖体，一个 mRNA 分子附着的核糖体的数目依赖于 mRNA 分子的长度和细胞系统。例如，血红蛋白合成系统中每个 mRNA 分子有 4 至 5 个核糖体附着(形成一个多聚核糖体)，而血清清蛋白系统，每个多聚核糖体约含有 15 个核糖体。在这二个系统中核糖体数目和 mRNA 分子中核苷酸数目的比率都为 1:80，也就是说，多聚核糖体大小的不同渊源于mRNA分子的长度。然而，在大肠杆菌的色氨酸系统，估计比例约为每 22 个核苷酸合一个核糖体。mRNA 分子的 22 个核苷酸长度接近于 150Å，而在电子显微镜下测量到的核糖体直径为 180Å，可见整个 mRNA 分子都被拥挤不堪的核糖体串满了。

一个多核糖体如果含有 5 到 15 个核糖体，即意味着有 5 到 15 个成熟程度不同的肽链同时在转译。我们不禁要问：是什么调节核糖体在mRNA分子上的附着，是什么使蛋白质合成终止和使核糖体脱离 mRNA。就多顺反子 mRNA 而论，第二个问题更使人感到兴趣。在那里，某一个的核糖体合成一个肽链之后，并不离开 mRNA，而是使另一个肽链开始转译。

在细菌中，蛋白质合成的起始明显地依赖于起始密码子 AUG

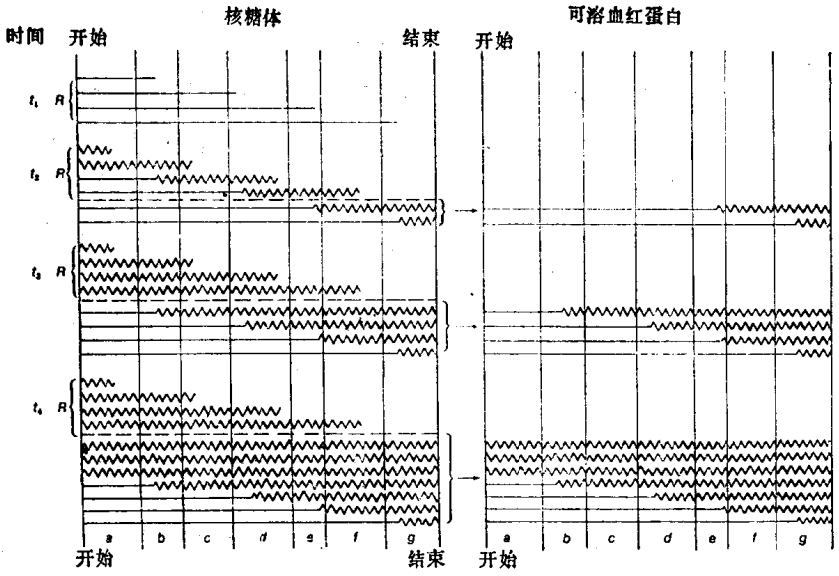


图 129 蛋白质的合成方向

这些实验包括在生长血红蛋白肽链的系统中加入 ^{14}C 标记的氨基酸，加进标记氨基酸的持续时间不同(t_1 到 t_4)，并在不同的时间测定标记首先出现在N末端呢，还是在C末端。在引入放射性氨基酸以后，把血红蛋白消化成片断肽，测定每段肽的放射性（已知这些片断肽在血红蛋白肽链上的次序）。图中曲线表示标记的片断肽，而直线表示无标记的片断肽。t_1 无标记肽链，但是新肽链处于不同的合成阶段，一些正在开始，另一些接近成熟。因此，在 t_2 时，^{14}C 能结合在肽链的开头、中间和末尾。如图所示，t_2 时标记在N末端的是开始合成的肽链，而将近成熟的肽链标记在C末端。(摘自H M. Dintzis, Biochemistry, 47, 1961。)

和GUG的作用，但是不知道这些密码子是否调节核糖体和mRNA的结合配置。如果回想一下 Nirenberg 的早期实验，就会想起在多聚 U、多聚A等模板上合成的蛋白质。这些模板没有密码子 AUG 和 GUG，然而却发生蛋白质合成和核糖体的配置。这个现象一直得不到解释，但最后发现在低浓度的 Mg^{2+} 溶液中，蛋白质合成不需要起动密码子。低浓度 Mg^{2+} 溶液可使蛋白质在不存在起始密码子的情况下开始合成，有人认为这是离体状态所产生的“错误”。甲酰甲硫氨酸和 $tRNA^{fMet}$ 的确切作用还不甚了解，而且似乎哺乳动物的蛋白质合成系统并不需要类似的起始因子（虽然线

粒体系统的蛋白质合成需要 AUG 和 $tRNA^{fMet}$)，而是由专一的但非甲酰化的甲硫氨酸-tRNA 起动的。

肽链终止机理的发现又另有一番经历。最初，Nirenberg发现64种密码子中的61种能识别氨基酸，只留下三个密码子 UAA、UAG和UGA与众不同。当时他曾建议把这三个密码子称作无意义密码子或者终止密码子。通过 H. Gobind Khorana 的巨大努力，才使制造具有预定核苷酸顺序的人工 mRNA 分子成为可能(图 130)，人工合成含有 UAA、UAG 和 UGA 密码子的 mRNA 后，才最终证实这三个密码子是终止密码子。

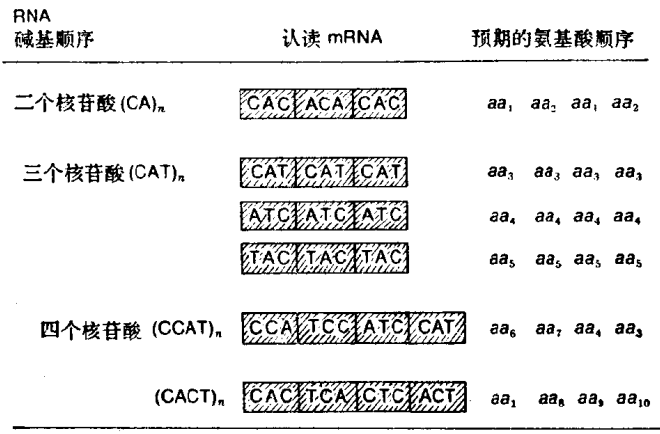


图 130 人工合成的 mRNA

以重复二个碱基开始，然后再重复三个碱基，Khorana 合成了各种预定顺序的mRNA，分析经人工合成mRNA分子转译成的蛋白质中氨基酸的顺序，鉴定了终止密码子的编码和功能。

细菌顺反子的突变能产生终止密码子。由于发现终止密码子离顺反子转录起始的距离和转译产生的多肽片断的长度有关，在这种情况下，能够用遗传学方法展示无意义突变的位置。可以想象，在这样的系统中，终止密码子在 mRNA 的位置，合成的多肽片段的长度和遗传图距三者之间必定密切相关(图 131)。

在转译过程中遇到终止密码子，50S 核糖体亚基的氨酰-tRNA 结合位置上就出现一个空白，这时肽链延伸停止，并且与

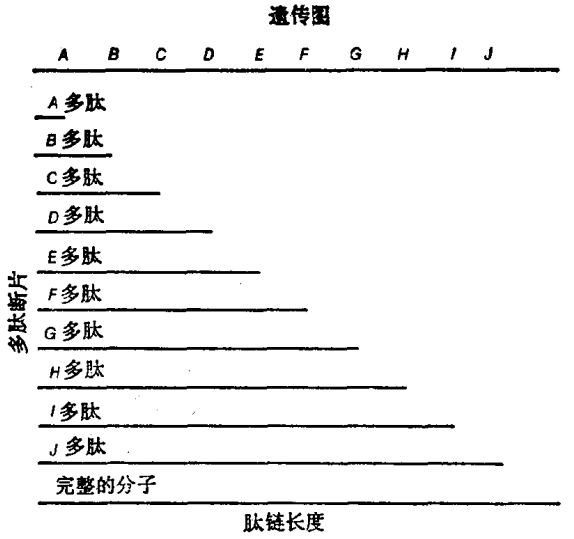


图 131 特殊的人工 mRNA 分子的合成

UAA、UAG和UGA密码子不和氨酰-tRNA结合，也无从定义它们的功能。有一个实验得出这些密码子是转译终止信号的结论。所用材料是 T_4噬菌体一些编码头部蛋白基因的突变型，这些突变型产生长度不同的不完整的蛋白质。图示终止密码子在基因中的位置(遗传图)和它们产生的蛋白质断片。(摘自 Brenner。)

核糖体脱离，也就是说在空白出现的时刻，氨酰-tRNA 结合位置的密码子是终止密码子。整个复合物的结构如下(图 132)：

fMet—Naa_1—aa_2—aa_3⋯aa_n—CO—tRNA—核糖体—mRNA

还不了解这个复合物的分离和个别单元再循环的详细过程。已经分离出一些所谓的释放因子，其中一个因子的功能被假定为释放fMet，另一个的功能是把氨酰-tRNA 和氨基酸相分离。还有二个功能尚未确定的因子，然而大致可以设想它们的作用是把作为蛋白质合成场所的复合物予以卸除，相当于装配复合体的逆过程。现在我们还不能解释所有步骤的详细化学变化，但即便如此，可以肯定蛋白质合成的程序图和蛋白质合成的实际过程不会相差太远。

了解蛋白质合成的一般程序之后，再仔细考虑一些先前介绍的，与其说是解答疑点不如说是罗列实例的遗传学现象是饶有兴味的。噬菌体 T_4 r_{II} 缺失突变型可算其中一例。这个突变型的顺

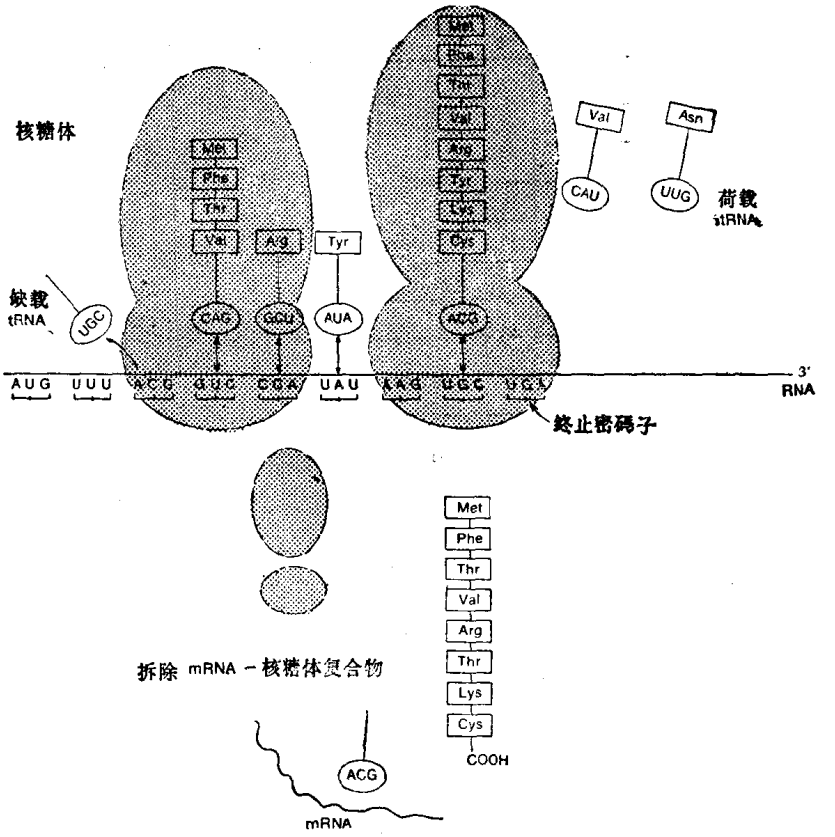


图 **132** 蛋白质合成期间的终止密码子

在mRNA分子的转译期间，缺载tRNA分子从肽基结合位(P)解离，而氨酰-tRNA分子被吸收到氨酰-tRNA结合位(A)，若进入氨酰-tRNA结合位(A)的是终止密码子，氨酰-tRNA就不再结合上去。但运载完毕的缺载tRNA仍在释放，由于没有新的tRNA移到肽基P位，由肽基转移酶催化的终止反应释放出不完整的多肽，余下的mRNA顺序得不到转译。

反子A和B相互连接，以至A顺反子的突变可引起B顺反子转录和转译的变化。二个读码节段的合并能简单地以这二个顺反子之间一个或几个终止密码子的丧失来解释。在这种情况下，r_{II}缺失突变型失去的片断包括一个或几个终止密码子，从而使二个顺反子的核苷酸联合成一个转录结构。在理论上，失去终止密码子必然使分别转录和转译的二个读码节段并成一个。

学科的进一步发展要求能提供大分子结构和功能模型与基础遗传学基本定律具有内在联系的实验证据。现在人们终于证明，三种大分子化合物之间所存在的线性关系（指线性排列次序相应的关系——译注），以及密码子的碱基改变，蛋白质中氨基酸的改变和遗传图之间的相关性等问题。Yanofsky 和他的同事为阐明这些关系提供了大量证据。

Yanofsky 等人致力于色氨酸合成酶的研究，在大肠杆菌中，色氨酸合成酶蛋白由二种不同的多肽即α多肽和β多肽组成。每条α肽链的大小为 29,000 道尔顿，每条β肽链为 50,000 道尔顿，有功能的酶由二条α链和二条β链装配而成。色氨酸合成酶催化的反应如图 133 所示。

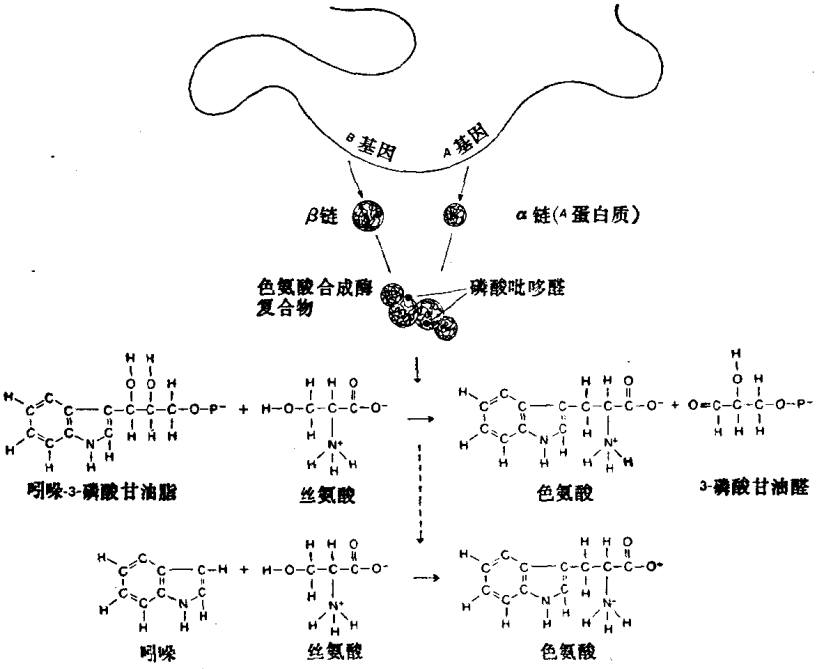
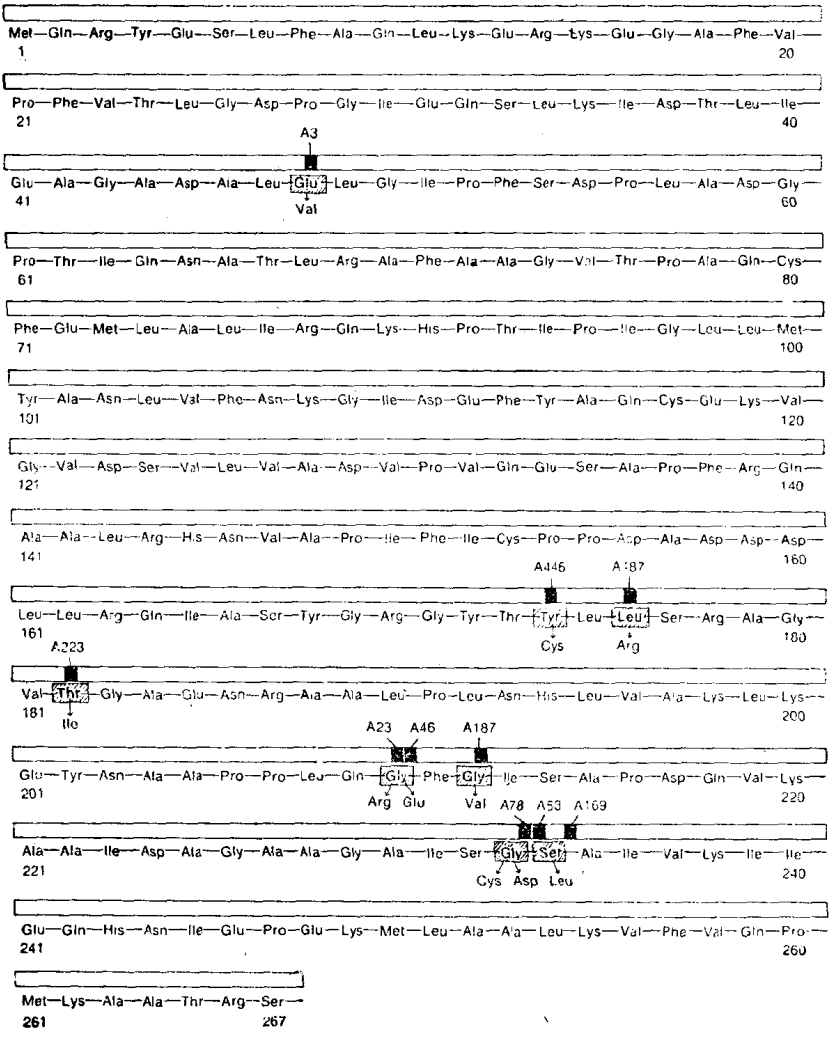


图 133　大肠杆菌色氨酸合成酶

图示色氨酸合成酶催化的反应，这个反应的正常生理作用是把吲哚-3-磷酸甘油酯和丝氨酸转化为色氨酸和 3－磷酸甘油醛。色氨酸合成酶由四条肽链组成，二条α链和二条β链。α链(A蛋白质)由A基因编码，β链由B基因编码。

(引自 C. Yanofsky, 1967。)


Met—Gln—Arg—Tyr—Glu—Ser—Leu—Phe—Ala—Gln—Leu—Lys—Glu—Arg—Lys—Glu—Gly—Ala—Phe—Val—
1
20
Pro—Phe—Val—Thr—Leu—Gly—Asp—Pro—Gly—Ile—Glu—Gln—Ser—Leu—Lys—Ile—Asp—Thr—Leu—Ile—
21
40
A3
Glu—Ala—Gly—Ala—Asp—Ala—Leu—Glu—Leu—Gly—Ile—Pro—Phe—Ser—Asp—Pro—Leu—Ala—Asp—Gly—
41
Val
60
Pro—Thr—Ile—Gln—Asn—Ala—Thr—Leu—Arg—Ala—Phe—Ala—Ala—Gly—Val—Thr—Pro—Ala—Gln—Cys—
61
80
Phe—Glu—Met—Leu—Ala—Leu—Ile—Arg—Gln—Lys—His—Pro—Thr—Ile—Pro—Ile—Gly—Leu—Leu—Met—
71
100
Tyr—Ala—Asn—Leu—Val—Phe—Asn—Lys—Gly—Ile—Asp—Glu—Phe—Tyr—Ala—Gln—Cys—Glu—Lys—Val—
101
120
Gly—Val—Asp—Ser—Val—Leu—Val—Ala—Asp—Val—Pro—Val—Gln—Glu—Ser—Ala—Pro—Phe—Arg—Gln—
121
140
Ala—Ala—Leu—Arg—His—Asn—Val—Ala—Pro—Ile—Phe—Ile—Cys—Pro—Pro—Asp—Ala—Asp—Asp—Asp—
141
160
A446
A187
Leu—Leu—Arg—Gln—Ile—Ala—Ser—Tyr—Gly—Arg—Gly—Tyr—Thr—Tyr—Leu—Leu—Ser—Arg—Ala—Gly—
161
Cys
Arg
180
A223
Val—Thr—Gly—Ala—Glu—Asn—Arg—Ala—Ala—Leu—Pro—Leu—Asn—His—Leu—Val—Ala—Lys—Leu—Lys—
181
Ile
200
A23
A46
A187
Glu—Tyr—Asn—Ala—Ala—Pro—Pro—Leu—Gln—Gly—Phe—Gly—Ile—Ser—Ala—Pro—Asp—Gln—Val—Lys—
201
Arg
Glu
Val
A78
A53
A169
220
Ala—Ala—Ile—Asp—Ala—Gly—Ala—Ala—Gly—Ala—Ile—Ser—Gly—Ser—Ala—Ile—Val—Lys—Ile—Ile—
221
Cys
Asp
Leu
240
Glu—Gln—His—Asn—Ile—Glu—Pro—Glu—Lys—Met—Leu—Ala—Ala—Leu—Lys—Val—Phe—Val—Gln—Pro—
241
260
Met—Lys—Ala—Ala—Thr—Arg—Ser—
261
267

(a.)

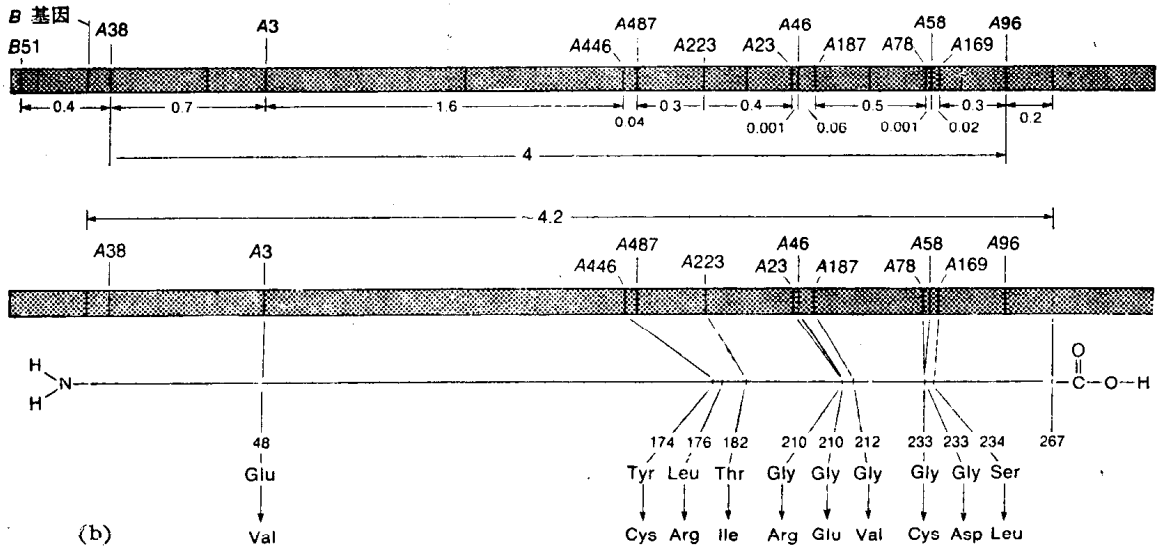
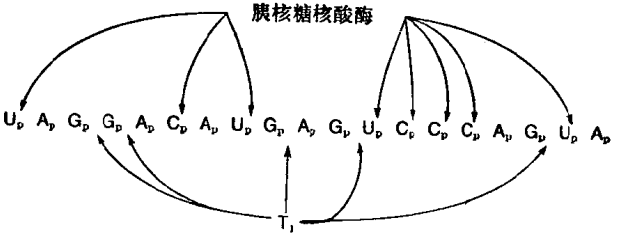


图 134　遗传图和氨基酸顺序的线性关系

比较遗传图和相应蛋白质的氨基酸顺序，能够测定基因和对应的蛋白质的线性关系。在A基因决定 A蛋白的系统中，突变定位采用普通的遗传学方法，突变型中置换的氨基酸可通过分析A蛋白质中的氨基酸顺序加以检测。A 蛋白质中的氨基酸顺序与A基因的各种突变表现有对应关系，遗传重组分析不能测定突变位点间的精确距离，但能证明突变次序和氨基酸被置换的次序相对一致。(引自 C. Yanofsky, 1967。)

编码α肽链的基因称为 *trpA* 基因，正是该 *A* 基因和α肽链，人们就其线性关系进行过深入的分析比较。曾经大量诱导了 *A* 基因各种类型的突变，包括碱基置换，终止突变，移码突变和各种缺失突变，发现碱基置换位于基因一定的区域，而终止突变随机地散布于整个基因中。

实验表明单位点突变特别适于当作标记，使用这种标记可把基因中发生的变化和α肽链中发生的变化进行比较。从图 134 上可以看到，DNA 的核苷酸变化和蛋白质中相应氨基酸的变化表现直线性相关(指直线排列次序相同)。早先曾假设核苷酸之间的距离和氨基酸残基的距离成比例，但是直到能够定出基因的核苷酸顺序之后，这个假定才得到验证。能够定出核苷酸顺序迄今已有数年，但工作本身仍十分困难(图 135)。从现有资料得知，遗传图距和氨基酸残基的距离之比并非完全一律，换言之，虽然基因的结构各部分是相当"均一"的，但整个基因组中各个重组位点发生重组的频率却明显不同。从位点 48 到位点 234 包含 186 个氨基酸残基的一段图距接近 3 个图距单位，即 0.015 个图距单位相当于一个氨基酸，也就是相当于 DNA 的三个核苷酸。现将较长的 DNA 划分成重组频率较高以及较低的片段时，计算出的图距和氨



(a)

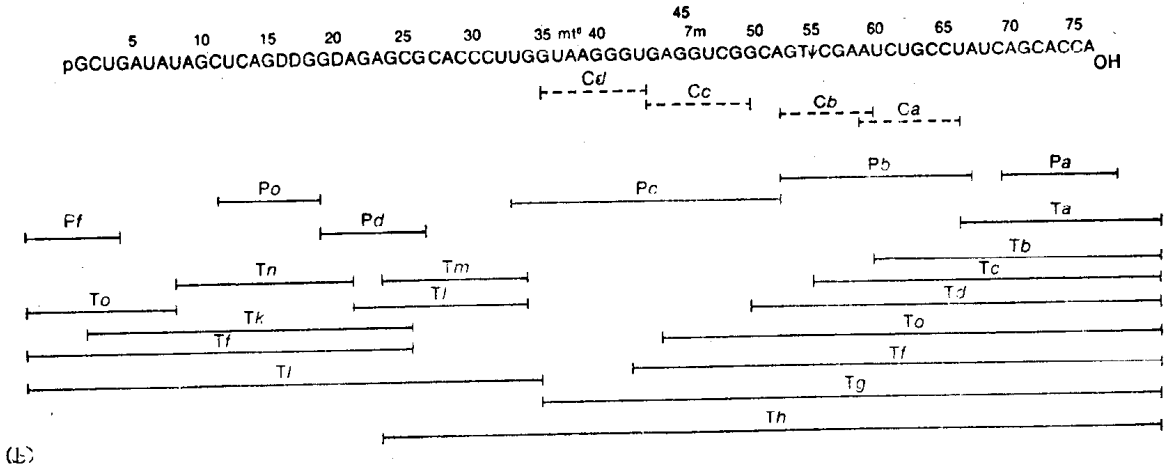
胰核糖核酸酶

U, AGGAC, AU, GAGU, C, C, C, AGU, A, U, C,

T_1核糖核酸酶

UAG, G, ACAUG, AG, UCCCAG, UA

} 低聚核苷酸断片和一些单核苷酸



(上)

图 135 列出DNA顺序

利用核酸酶能列出DNA的顺序。某些核酸酶切开多聚核苷酸有专门的位点。例如，仅在胞嘧啶核苷酸的位置或者仅在鸟嘌呤核苷酸和胞嘧啶核苷酸之间等等。如此某个核酸酶可切出许多类型特殊的断片(低聚核苷酸)，用层析法分离断片，然后用外切核酸酶定出每个断片的顺序，外切核酸酶从低聚核苷酸的一端开始，每次切下一个核苷酸残基。

(a) 分批处理DNA，每批DNA分子用不同的核酸酶处理，都可产生二组或二组以上的断片。

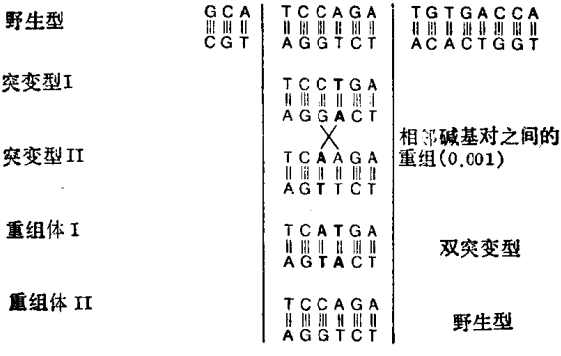
(b) 列出从二种或二种以上酶解反应中得到的低聚核苷酸断片的顺序，可以看到这些顺序交互重迭，利用重迭顺序可确定低聚核苷酸断片在DNA分子上的排列次序。图示苏氨酸tRNA的顺序。(引自 L. Clarke与 J. Carbon。)

基酸残基的比率分布在 0.01 到 0.05 之间，高低相差 5 倍之多。如果能够比较与每两个氨基酸对应的密码子之间的重组频率，那么这个比率差距可能更大。

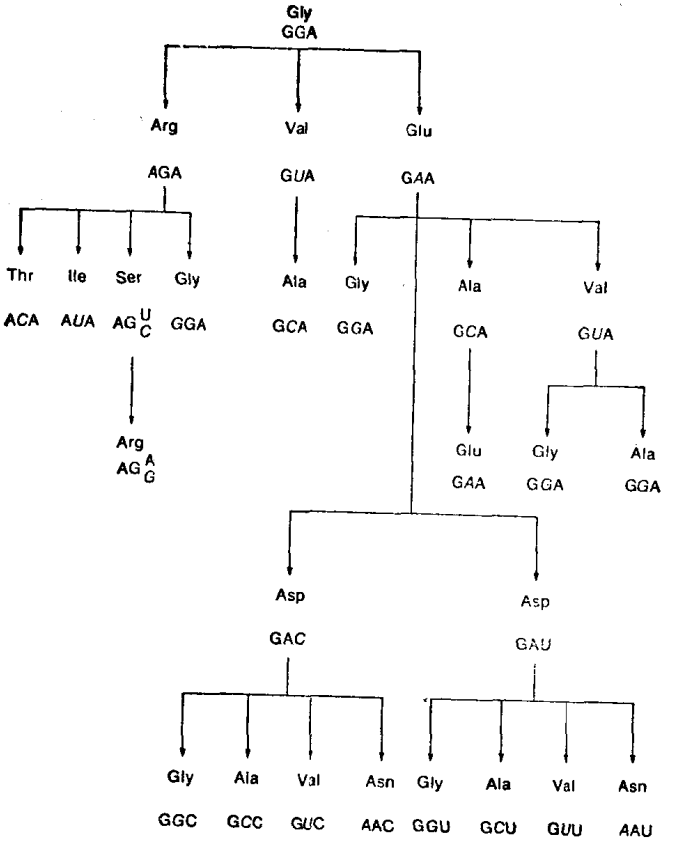
观察一些发生在同一个密码子内的各种类型的突变尤为有趣，各个突变型产生的 α 肽链在某个位置的氨基酸互不相同。这些突变型包括单个密码子内核苷酸对的碱基置换。观察到密码子内的重组频率(相邻核苷酸对之间的交换值)为 0.001，这个数值大大低于预期值 (0.005 或 $\frac{1}{3}\times 0.015$)，即低于基于总的图距和肽链长度的计算数值(图 136)。

曾对各个突变型，包括回复突变和抑制基因突变，与相应的发生在 α 肽链第 210 位上的氨基酸替换作过对比(图 136)，总共发现 19 种氨基酸替换和 17 种罕见的密码子。许多突变型多肽是有功能的，而另一些则无功能。当肽链的第 210 位为精氨酸、谷氨酸或天冬酰胺时，色氨酸合成酶全无催化活性。若为另一些氨基酸，则表现从低到高一定范围内的酶活性（野生型蛋白质的第 210 位是甘氨酸)。从 A 基因中不同位点突变与 A 蛋白质中氨基酸替代的分析，可据而比较密码子和氨基酸在活体内的对应关系，最终证实 Nirenberg 和 Leder 在体外实验中提出的指配关系是正确的。

在分离适用的突变型菌株和发展 A 基因-α 肽链系统以分析



(a)



(b)

图 136　密码子内的重组

(a) 相邻核苷酸对(在一个密码子内或在密码子之间)的重组例图。观察到相邻核苷酸之间的重组频率为 0.001，这个数值大大低于从比较图距和肽链长度算得的预期值。例如，据计算单个氨基酸残基占有 0.015 个图距单位，一个氨基酸的间隔和三个核苷酸对相等。因而每个核苷酸对的图距应为 0.005 个单位，这个数值是观察值的 5 倍。

(b) 分析一定数量的突变和回复突变，发现A蛋白质第 210 位共有 19 种不同的氨基酸。例图从识别甘氨酸的野生型密码子GGA开始，列出可能在突变型和回复突变型的大肠杆菌菌株中编码氨基酸的密码子。(引自C. Yanofsky, JAMA 218:1026, 1971。)

线性关系的过程中，Yanofsky 把许多称为增变基因 (mutator gene)的突变菌株进行杂交。过去在某些物种如玉米和果蝇中曾讲到过这种增变基因，但是以前从未把它们作为在 DNA 和蛋白质

水平上进行分析的材料。大肠杆菌的增变基因可把某些位点的突变频率从 10^{-8}提高到 10^{-5}。在受增变基因影响的位点中有5个编码与色氨酸合成途径有关的酶的结构基因，其中包括 *trp*A 基因。分析 *trp*A 基因内增变基因诱导的氨基酸替换，发现突变类型不是随机的，而是 AT→CG 颠换首先被诱导产生(图 137)。

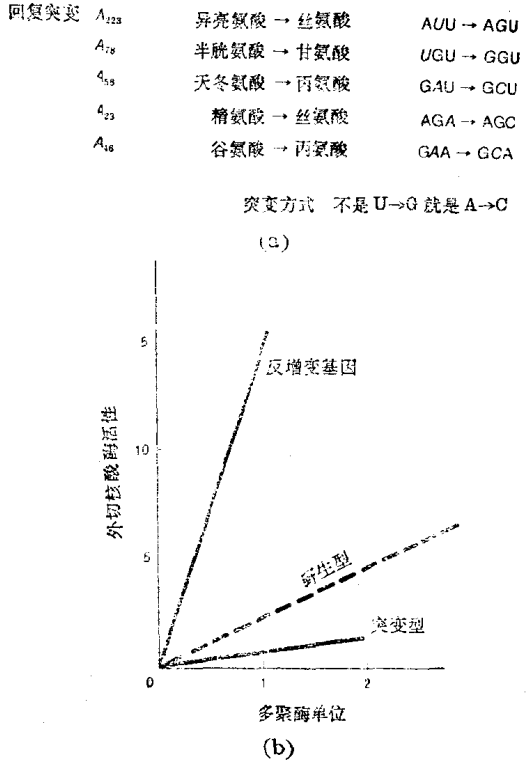


图 137 增变基因

增加其他基因突变频率的增变基因，在许多物种内都有存在，过去从不明了它们的作用方式。但是 Yanofsky 实验室观察了色氨酸合成酶的 A 蛋白的变化，至少对于理解某一类增变基因的功能有重大意义。

(a) 发现许多位点(图示五个)的回复突变不是U→G 就是 A→C，从而表明了增变基因对其他突变的效应是非随机的 (引自 Yanofsky et al., proc Natl Acad Sci USA, 55:274, 1966。)

(b) 研究了另一类变更3′→5′外切核酸酶活性的增变基因，一个这样的突变降低外切核酸酶的活性，而另一个则提高了酶活性。提高外切核酸酶活性的菌株具有降低天然突变的性能，被称为反增变基因 (antimutator)。(引自 Kornberg, 1974。)

增变基因在染色体许多位置上优先诱导碱基对颠换，导致一个物种的 DNA 存在大不一样的 AT 与 GC 比例，从而引起了引人入胜的有关进化趋势(evolutionary trends)的问题。例如，密码简并准许变动许多密码子的第三位核苷酸，而不影响对应蛋白质的氨基酸，也就是说 AT 和 CG 的比率能够在有限的范围中波动而不改变蛋白质的氨基酸成分。加之，许多蛋白质能经受某些氨基酸组分的替换而不失去酶活性，亦即允许可能存在有多种多样的 AT 对 GC 的比率。

曾经连续好几代对带有增变基因或不带有此基因的大肠杆菌菌株，在细胞分裂期间就 AT 与 GC 对比率的变化进行过比较，发现在世代进化过程中，带有增变基因的菌株的 DNA 增加了 GC 对而减少了 AT 对。既然 GC 对含量较高的 DNA 的浮力密度较大，故能使用这一方法来检测 DNA。

分析一定数量的细菌和噬菌体系统的增变基因后，得出了这样的规则，这些增变基因能够影响染色体组任何部位的突变。但是每种增变基因诱发的突变种类则有其特殊性。一些增变基因专门作用于 DNA 的一条链，而另一些却能够诱发移码突变。T_4 噬菌体至少有一种增变基因能变更 DNA 多聚酶活性，随后在 DNA 复制期间又据而变更野生表型的“差错幅度”(mistake level)。在某些场合差错幅度反而减少了，换言之，增变基因还具有产生某种反增变多聚酶的作用。

蛋白质合成、各种 RNA 在合成期间的功能、核苷酸顺序转译成氨基酸顺序以及遗传密码等包括了大量复杂的生化反应。但是这些反应都能够将之概括于遗传信息流的总过程之内（图 138)。从实验的观点来看，把信息流程中的诸种生化反应的脉络展示出来，会大大有助于认识那些同样复杂的与蛋白质合成有关的细胞学过程。限于篇幅，这方面的大量内容不能在此介绍。

我们的第一个目的是提供整个生物信息传递过程的简明梗概。下一步将考虑蛋白质。由转译而承受了遗传信息的蛋白质，它们的功能是促进表型的发育。一旦由 DNA 经过 RNA 合成了

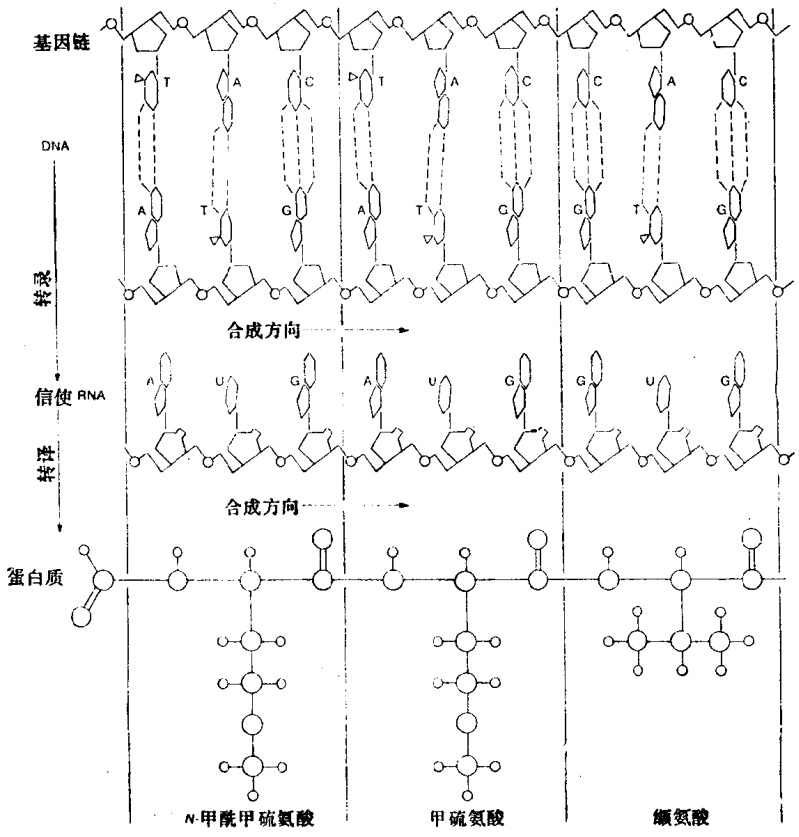


图 138 信息从 DNA 传递到蛋白质

从大分子本身的有利特点，可知遗传信息系储存于DNA分子的核苷酸排列顺序之中，以后转录为mRNA的核苷酸顺序，再转译为多肽的氨基酸顺序。也正是按照这一特征才有可能设想多肽的信息，转变为表现型的图景。为此，第一步应分析一下多肽和蛋白质的结构与功能。(引自Clark 与 Marcker, 1968。)

蛋白质，它们行使功能一般独立于产生它们的诸种蛋白质合成器。然而，如我们将看到的那样，某些蛋白质的作用是参与复制、转录和翻译等过程的调节，有关蛋白质的特性和功能问题将在下一章讨论。

参 考 文 献

Cold Spring Harbor Symposium of Quantitative Biology: *The Mechanism of Protein Synthesis*, vol. XXXV, 1969.

Sirlin, J. L.: *Biology of RNA*, Academic, New York, 1972.

Woese, C. R.: *The Genetic Code*, Harper & Row, New York, 1967.

Ycas, M : The Biological Code, in A. Neuberger and E. L. Tatum (eds.), *Frontiers of Biology*, vol. 12, North-Holland, Amsterdam, 1969.

第五章　蛋白质：遗传信息转化为表现型的过程

深刻揭示核酸在结构和功能之间所存在的内在统一性是分子遗传学研究中举世瞩目的重大突破之一，它雄辩地指出复制和转录机制在其总的体制上是与 DNA 和 RNA 分子的主要结构特征紧密相联的。同样，转译的结构与功能之间相互关系的模式也可以从蛋白质合成的一般知识以及蛋白质据以合成的大分子复合体的性能加以说明。但是，这里所涉及的生化机制，蛋白质合成装置的结构以及各种 RNA 的协同作用却不能从 RNA 的结构中得出明确的结论。至于它们同蛋白质结构的关系就更加模糊不清了。

蛋白质的组成元件本质上不同于核酸的字母表，但是蛋白质的一级结构（氨基酸残基的排列顺序）却是由核酸的一级结构（核苷酸的排列顺序）所决定的。这二种类型的大分子化合物都是由种类为数不多的单位元件连接而成的多聚长链。当然，这仅仅是对二者进行大致比较的情形而言。蛋白质与核酸不同，它不能将其包含在一级结构中的遗传信息传递给其他的大分子，例如，迄今为止尚未发现有依赖于蛋白质的 RNA 聚合酶。蛋白质乃是以另一种极为不同的方式来表达它们的遗传信息，其中某些方式非常复杂，根本不易从它们的结构本身来分析推断。

蛋白质在细胞和生物体中具有多种功能，但是我们在这里主要着重于讨论它的催化功能。同新陈代谢、生长、分化以及生殖有关的生化反应一般都由酶来催化，大多数酶都对单个反应或某一类有关反应具有专一性。这些反应从电子由一种蛋白质转移到另一种蛋白质起，直到聚合百万个以上核苷酸长度的 DNA 分子，类型繁多，范围极广。目前，还缺少任何可以根据蛋白质的物理性

质来对酶分子功能作出预测的切实可行的方法。在蛋白质的多种性质中，最重要者应数蛋白质分子上氨基酸的排列顺序。尽管有许多种蛋白质的氨基酸顺序已经敲定，但是关于氨基酸顺序怎样转变为功能上的特异性问题目前还是不得而知。

至于遗传信息流在蛋白质一级结构水平以上如何传递表达，我们几乎一无所知，这一知识上的空白并不妨碍我们说明蛋白质的自然性质。但是，就蛋白质的自然性质和功能之间的关系来说，由于人们对这一自然规律的认识尚属浅薄无知，结果往往会使分子遗传学的初学者们对于许多可能与遗传信息流的一般理论不尽符合的事实和材料感到困惑不解。

氨 基 酸

整个自然界有数百万种蛋白质，但是作为基本结构单元的氨基酸一共只有 20 种，因此，它们之间在一级结构上的差异仅仅在于这共同的 20 种氨基酸在排列长度和排列顺序上互不相同而已。并非所有蛋白质都包揽无遗地含有这全部的 20 种氨基酸，甚至在某些罕见的蛋白质中还发现有一种或几种例外的稀有氨基酸，但是从蛋白质一级结构的一般组成来看，作为蛋白质结构单元的氨基酸就不过这 20 种。图 139 所列即为这 20 种氨基酸的粗略分类。每一种氨基酸的R基团，即侧链都各不相同，其共同特点是均有一个 α 碳原子(图 140)，一个 α 羧基，一个 α 氨基和一个连接在 α 碳原子上的侧链。各个氨基酸分子彼此以羧基和氨基首尾相连缩合而成肽键。因此，在多肽链的氨基酸残基之间肽键起有契合作用。侧链的差异一方面说明了氨基酸残基之间在化学性质上的区别，另一方面，在某种程度上也能够解释蛋白质之间理化性质上的特点。

根据氨基酸侧链的结构特点和化学性质，可将氨基酸分成如下 8 类：1) 酸性氨基酸(带有羧基的酸性侧链)；2) 碱性氨基酸(带有氨基的碱性侧链)；3) 脂肪族氨基酸(带有直链的侧链)；4) 羟基氨基酸(含有羟基的侧链)；5) 芳香族氨基酸(带有芳香环的

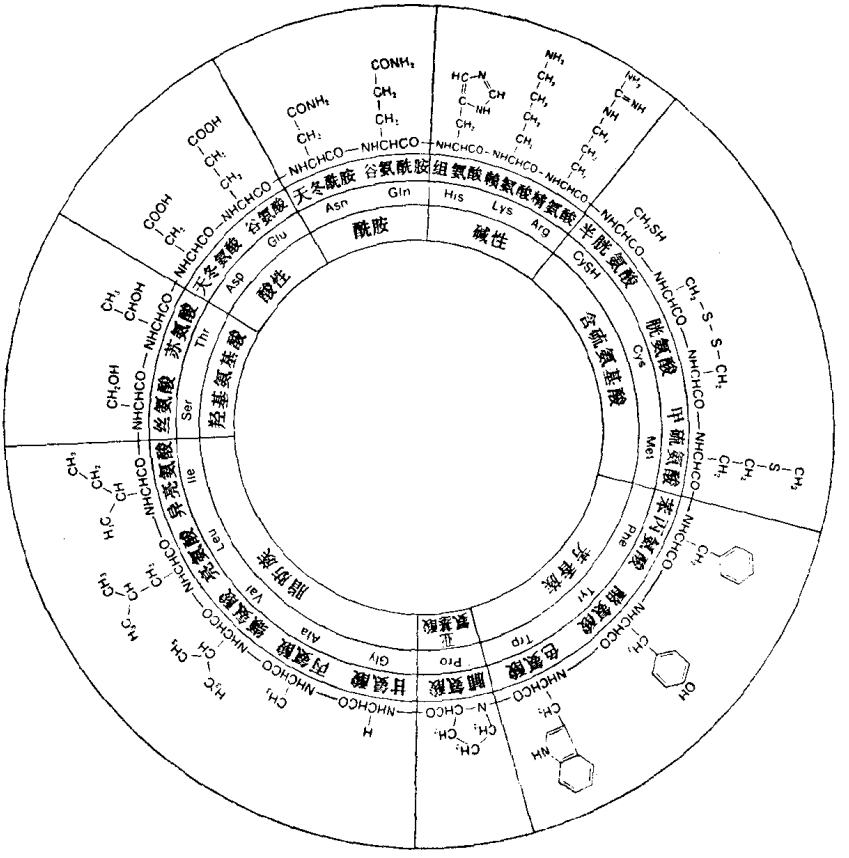


图 139　氨基酸

二十种最常见的氨基酸化学结构，分别按其侧链的化学性质和结构进行归类。(引自 Alfred A. light, Proteins Structure and Function, Prentice-Hall, 1976。)

侧链)；6) 含硫氨基酸(带有硫氢基侧链)；7) 酰胺类；8) 亚氨基酸类，这种氨基酸的特点是 α-碳原子上的侧链又和α-氨基相连(图 139)。有时同一个氨基酸可以适合上述分类标准的几个门类，例如，酪氨酸的侧链就同时含有芳香环和一个羟基。

除甘氨酸外，所有氨基酸都具有不对称性 α 碳原子(图 140 表示 α 位碳原子与 4 个不同的基团连接)。这意味着对于每个不对称碳原子都可能存在有二种结构构型，通过分子旋光度分析法可

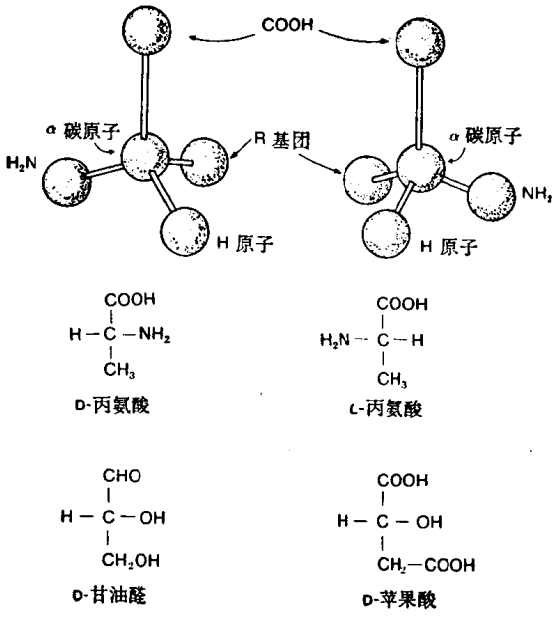


图 140　氨基酸的不对称性和分子构型

所有氨基酸(除甘氨酸外)的α碳原子都具不对称性,即与其连接的四个基团在化学性质和结构上都不相同。一个为氢原子,一个为氨基,一个为羧基,第四个是R基团(R代表侧链,这个基团对各种氨基酸都是不同的)。

不对称性氨基酸存在二种构型,均可通过氨基酸偏转偏振光的方向来确定。通常D型表示氢原子在左,L型表示氢原子在右。图中间所示为D型丙氨酸和L型丙氨酸,下行所列为另外二种重要的分子:D-甘油醛和D-苹果酸,以资比较。

以检测出两种构型的差别。由于甘氨酸侧链是一个氢原子,因此甘氨酸的α碳原子连接的就是二个氢原子。它的构型比较特殊,呈对称性,也就是不具旋光性。除甘氨酸外,其余各种氨基酸均属于L型和D型这二种旋光异构体中之一种。如上所说,这可通过测量每种氨基酸旋转偏振光的方向,再以已知构型的甘油醛作为对照,就可以确定。

在天然蛋白质中迄今所发现的氨基酸大多数都为L型。D型氨基酸则属凤毛鳞角,一般只存在于某些细菌的细胞壁和低肽抗菌素中,如短杆菌肽和放线菌素等等(图141)。关于L型氨基酸

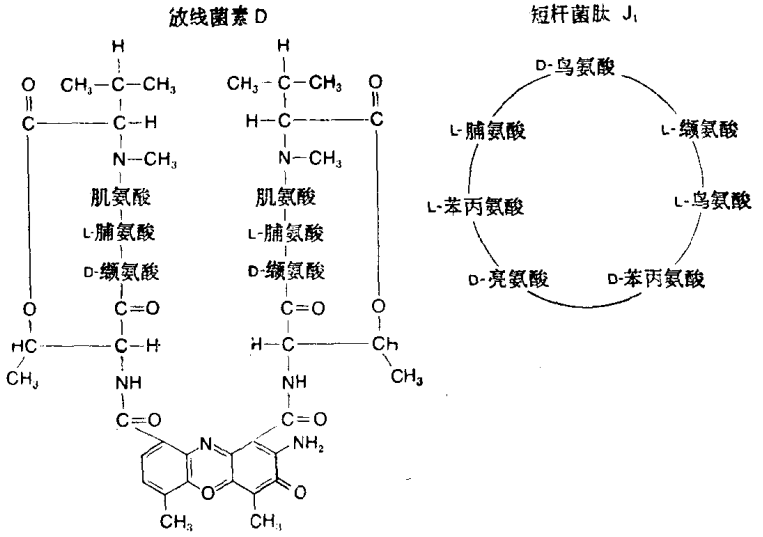


图 141　天然蛋白质中极为罕见的D型氨基酸

放线菌素D和短杆菌肽 J_1为二种含有D型氨基酸的小肽抗菌素。自然界中大多数的氨基酸均为L型旋光异构体。

的生物学意义是否一定就比D型氨基酸更为重要，长期以来一直争论不决。事实上现在仍然说不清楚在大多数天然蛋白质中肯定能够排除D型氨基酸的存在。

在 pH 值为 4～9 的溶液中，氨基酸以两性离子存在。这是因为每个氨基酸至少含有一个酸性基团 (COO^-) 和一个碱性基团 (NH_3^+)，亦即它们既具有羧基又具有氨基。羧基和氨基的电离反应表示为：

$$R-\overset{\overset{O}{\|}}{C}-OH \rightleftharpoons H^+ + R-\overset{\overset{O}{\|}}{C}-O^- \qquad (1)$$

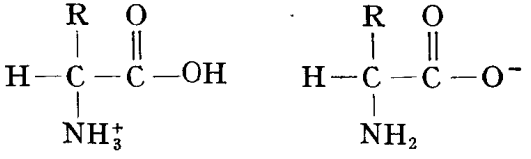
$$R-NH_3^+ \rightleftharpoons H^+ + R-NH_2 \qquad (2)$$

在任一时刻离子态和非离子态氨基酸的相对数量由溶液的 pH 值决定。电离常数K按下式计算：

$$K=\frac{[H^+]\times[RCOO^-]}{[RCOOH]} \quad 和 \quad K=\frac{[H^+]\times[RNH_2]}{[RNH_3^+]}$$

当氢离子浓度等于K值时，$RCOO^-$ 与 RCOOH 之比为 1，此刻的 pH 值称为 pK，或者叫做电离基团处于一半电离一半不电离时的 pH 值。因此 pK 即为氨基酸处于一半电离时的氢离子浓度。这个数值常被用来作为某个酸或某个碱的强度指标，并能行之有效地据以估算某一电离反应进行的程度。表 18 列出了几种氨基酸的 pK 值，这些数值常用于鉴定蛋白质中某些特定位置上的氨基酸。

大多数氨基酸在 pH 值低时带正电，在 pH 值高时带负电：



低 pH　　高 pH

由此可见，在某些处于中值 pH 的溶液中将会出现有数量相等的

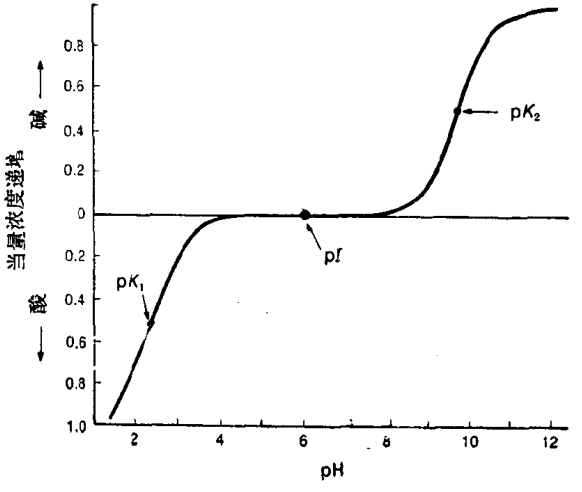
表 18　几种氨基酸的 pK 值

氨　基　酸	pK_1	pK_2	pK_3
精 氨 酸	1.914	9.718	12.48
天冬氨酸	2.1	10.6	3.9
半胱氨酸	1.7	10.8	8.3
谷 氨 酸	2.186	9.4	4.3
甘 氨 酸	2.35	9.8	—
组 氨 酸	2.4	10.8	6.0
赖 氨 酸	2.18	8.95	10.7
脯 氨 酸	2.011	11.296	—
丝 氨 酸	2.296	9.880	—
色 氨 酸	2.46	9.41	—
酪 氨 酸	2.2	10.05	9.1

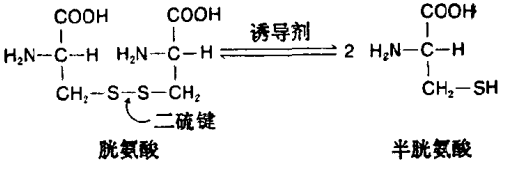
pK_1=羧基；　pK_2=氨基；　pK_3=侧链。

带负电荷的羧基和带正电荷的氨基。一般把这种状态下的 pH 值叫做等电点 pI，它依每种氨基酸分子的特性而异。等电点时，分子上无净电荷，分子在电场中也不移位。因为某些氨基酸具有一个以上的羧基(酸性氨基酸)，而另一些具有一个以上的氨基(碱性氨基酸)，因此蛋白质的等电点很大程度上取决于其一级结构中酸性氨基酸同碱性氨基酸的比例；比例高者，等电点便低。

图 142 例举的几个例子是描述有关氨基酸侧链的化学性质，



(a)

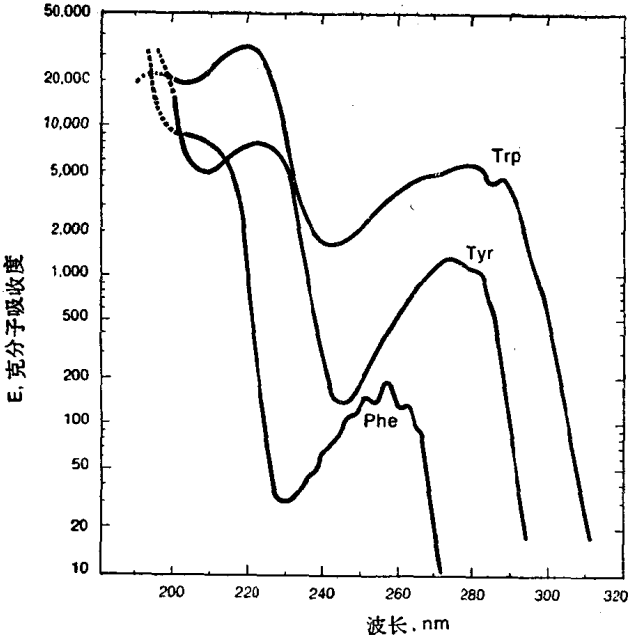


(b)

图 142　氨基酸侧链的若干化学性质

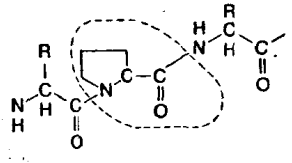
(a) 一氨基一羧基氨基酸的滴定曲线。氨基的pK值(电离基团-半电离-半不电离时的pH值)约为 9.6，羧基约为 2.3。这些数值采用滴定曲线(滴加酸或碱计算其克当量并记录产生的pH值变化)的方法获得。等电点pI为氨基酸分子处于电中性时的pH值。

(b) 二硫键。胱氨酸是一个二聚体分子，由二个以二硫键相互连接在一起的半胱氨酸分子组成。位于多肽链不同位置上的半胱氨酸之间的二硫键为蛋白质的折迭结构提供了基础。二硫键虽是共价键，但甚易被诱导剂切断。



(c) 亚氨基酸

脯氨酸　　羟基脯氨酸



一个多肽中的脯氨酸

(d)

图 142

(c) 图示紫外光吸收光谱。色氨酸、酪氨酸和苯丙氨酸的吸收光谱在 pH8的溶液中波长范围从 240nm 到 300nm。蛋白质的紫外光吸收光谱主要是依色氨酸和酪氨酸的光谱吸收性质而变化。

(d) 亚氨基酸。脯氨酸的碱性基团是亚氨基，而不象其他氨基酸那样一律为氨基。脯氨酸的氨基参与了一个五员环结构(吡咯R基团)，使其成为一个亚氨基。当脯氨酸参与形成肽键时，生成的键会制约多肽链的构象，而游离氨基一般地不会影响多肽链的构象。

以及如何从这些侧链的化学性质来鉴别蛋白质并说明它们与蛋白质化学性质之间内在的关系。下面可以看到，蛋白质最重要的理化性质之一是：作为一个有功能的蛋白质分子，它必须具有某种近似球状的由多肽链前后反复折迭而形成的特定构型（对大多数蛋白质而言），我们一般将这种构型称为蛋白质的三级结构，即三维结构。

功能蛋白质所具有的固定形状，部分是通过高活性的硫氢基氧化而使半胱氨酸侧链相互连接形成的二硫键来维持的（图142）；二个半胱氨酸分子侧链上的硫氢基，氧化成为二硫键之后便形成一个分子的胱氨酸。因此，胱氨酸不同于半胱氨酸，它在蛋白质合成时并不组合进蛋白质的一级结构，只是当多肽链一级结构顺序上位于不同位置的二个半胱氨酸残基由于肽链折迭而相互靠拢到足以形成一个二硫键时，胱氨酸才会脱颖而出。甲硫氨酸是侧链上含有硫原子的另一个无独有偶的氨基酸，但它从不形成二硫键，这一点颇与半胱氨酸不同。

由于芳香族氨基酸容易通过其紫外光吸收光谱加以检测，因而对蛋白质鉴定具有重要意义（图 142）。在中性 pH 条件下，色氨酸的吸收峰在波长 279 nm，酪氨酸的吸收峰在 275 nm，苯丙氨酸则在 258 nm。

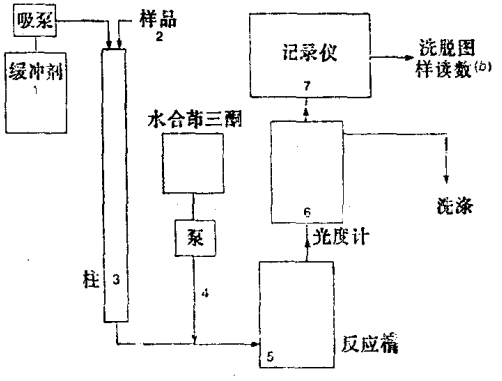
在生理 pH 范围之外（pH 4 以下和 pH 9 以上），酸性氨基酸和碱性氨基酸将会改变它们的电离状态，但组氨酸电离态的改变却在pH 4 和pH 9 之间。如低于 pH 4，组氨酸的咪唑侧链将增加一个质子，结果使侧链带上一个正电荷。在同组氨酸咪唑基 pK值相等的 pH 条件（即 pH＝6）下，有 50％的侧链带正电，50％为电中性。超过pH 9 组氨酸分子则表现为电中性。

脯氨酸不属 α-氨基酸而为一个亚氨基酸。由于所有 α-氨基酸在与显色剂水合茚三酮作用时都显示特异性的蓝色反应，唯独脯氨酸呈黄色反应，因此可与其他的氨基酸相区别。脯氨酸“侧链”上的 α 碳原子连在主链上亚氨基的氮原子上，结果在一个多肽链的各个脯氨酸位置上必定会显示出某种特征性的扭曲（图

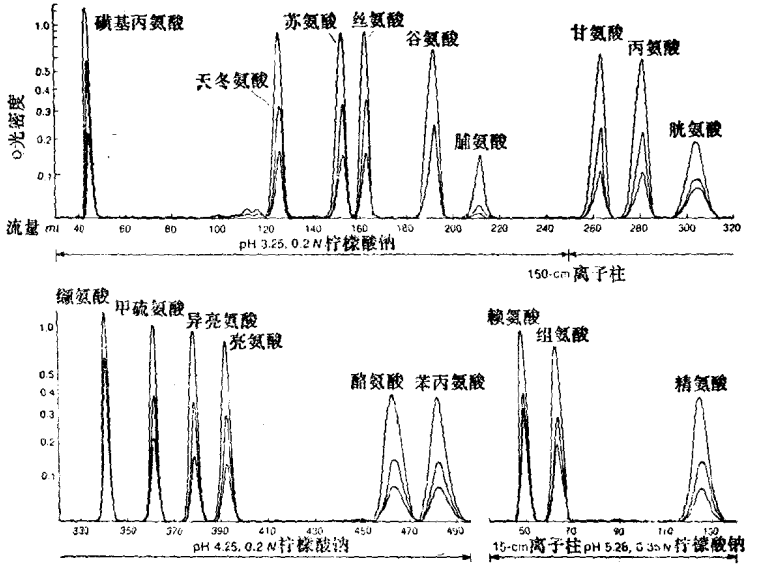
142)。由脯氨酸残基本身造成的多肽链结构上的这种扭曲对于蛋白质分子三级结构的形成有很重要的意义。

氨基酸侧链上—OH 和—NH 基团的质子能同其他的极性基团，例如多肽主链上羰基氧原子或组氨酸侧链环上的氮原子(在大于 pH 6 时)形成氢键。丝氨酸、苏氨酸、酪氨酸、色氨酸和组氨酸侧链也常常参与其事。另外，在多肽主链上的酰胺(—NH)质子和主链上的羰基氧原子之间也会形成氢键。有关氢键在蛋白质二级和三级结构中的作用我们将在后面讨论。

现今，氨基酸的鉴别和定量工作大部分已实现自动化。分析蛋白质化学特性的重要设备之一是已研制成功的称为氨基酸自动分析仪的装置（图 143)。将单一的或混合的氨基酸样品放入分析仪，使之通过一种特殊的树脂进行洗脱，这种树脂能根据氨基酸携带电荷的多少而将它们各自析离。经树脂洗脱析离后的氨基酸又流进一个盛有水合茚三酮的反应槽，水合茚三酮可将 α-氨基酸显为蓝色而使脯氨酸显为黄色。氨基酸在显色之后再流经一个比色计，这时已经显色的氨基酸即被自动记录成一种洗脱曲线（图 143)。在洗脱曲线中，峰值出现的顺序系取决于不同种类氨基酸各自流经树脂的洗脱速度。曲线峰内的面积则相应于氨基酸的数量。只要将已知种类和数量的氨基酸作好标准曲线，用来同未知



(a)



(b)

图 143 氨基酸分析仪

(a) 氨基酸分析仪的主体是离子交换树脂柱(3)和光度计(6)。所有其他的部件都是设计用于自动地和反复地进行氨基酸的洗脱与计量。磺酸聚苯乙烯交换柱(3)需保持恒温恒压。氨基酸样品(2)置于离子柱，由于每种氨基酸同树脂各有其特殊的亲和力，因而每种氨基酸就会以不同的速率从树脂柱底部被洗脱出来。不同的缓冲液(1)只能洗脱与之相应的氨基酸。pH3.25 的缓冲液可用来洗脱大多数的氨基酸：天冬氨酸，苏氨酸，丝氨酸，谷氨酸，脯氨酸和半胱氨酸(磺基丙氨酸)等；pH 4.25 的缓冲液洗脱甘氨酸，丙氨酸，胱氨酸，缬氨酸，甲硫氨酸，亮氨酸，异亮氨酸，酪氨酸，苯丙氨酸；而pH 5.28 的缓冲液则洗脱大多数碱性氨基酸：赖氨酸，组氨酸和精氨酸(见b下)。氨基酸从离子柱上洗脱后，又同水合茚三酮(4)混合，在一个加热槽(5) 中进行显色反应。然后将已同水合茚三酮反应的氨基酸通过光度计(6)，测定其特异性吸收光谱并自动读出洗脱曲线(b)。

(b) 图中氨基酸峰谱称洗脱曲线模型。当 pH、压力和流速一定时，洗脱峰是明显的。峰内面积相应于置入交换柱的氨基酸分子数目（重叠峰代表三次实验，每次采用不同的氨基酸浓度)。丝氨酸和苏氨酸在用酸水解蛋白质的过程中已部分降解，色氨酸也完全受到破坏，故在洗脱峰谱中看不到这三种氨基酸。

的混合氨基酸比较，即可对后者进行检测。其他方面的测试装置也已相当完善。例如有一种称为氨基酸顺序分析仪的装置，这一装置每次可处理一种完整的蛋白质或者多肽片段，它能够自动地从一段多肽中每次切下一个氨基酸，从而可以清楚地揭示出一个多肽链中全部氨基酸的排列顺序。

蛋白质结构

一级结构

具有催化功能的蛋白质分子存在着多种外部形态（或者说是

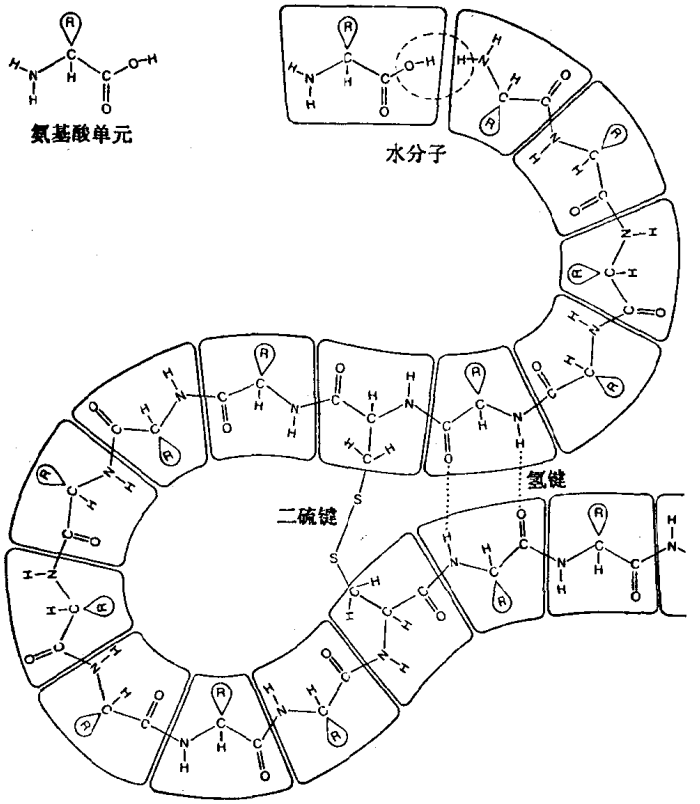


图 144 肽键

肽键在氨基酸的氨基和羧基之间形成。脱水后产生肽(酰胺)键。(引自 Dickerson与 Geis, 1969。)

状态)。在完全伸展的状态，即常常称为变性状态中，蛋白质显示为由氨基酸残基通过肽键相互连接而成的长链(图 144)。蛋白质合成时，肽链的生长同一个氨基酸的氨基与另一个氨基酸的羧基之间形成共价键有关。在这一反应中，一般是借助于产生水的过程使从羧基上脱去一个羟基，从氨基上脱出一个氢原子(图 144)。

多肽链的主链比较牢固。这种刚性产生于肽键是由部分双键组成的缘故(图 145)，使得肽键不能绕链轴旋转。大多数氨基酸的侧链基本上都可自由旋转，结果虽然主链比较牢固，但未经折迭的多肽链仍然是不稳定的分子。但无论如何，人们对未折叠多肽链最感兴趣的特性还是氨基酸单元的排列顺序。

实际上，蛋白质的性质取决于氨基酸的组成和氨基酸的顺序两大因素。对某一个多肽链来说，即使各个氨基酸残基的数目能精确地测定，但是搞清楚氨基酸的组成并不等于已经阐明了蛋白质的特性差异。这一指标只不过是对蛋白质之间差异的一个粗略的估计尺度，这是因为二种不同的蛋白质完全有可能具有相同的或颇相一致的氨基酸组成，但表现为不同的氨基酸排列顺序和不同的生理功能。反之，具有同一功能的蛋白质其氨基酸组成和排列顺序也可以相差悬殊。(在进化一章中我们将详细讨论有关细胞色素 c 蛋白的问题，细胞色素 c 蛋白可以从为数甚多的生物体中分离获得，它们都有着同样的功能，但却具有不同的氨基酸组

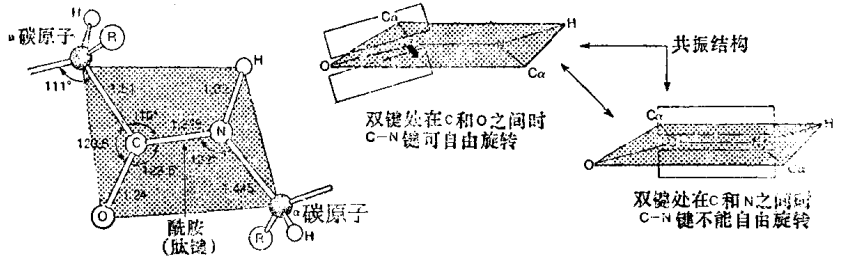
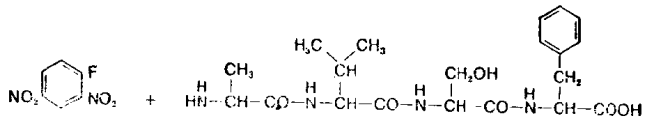


图 145　肽链中酰胺键上部份双键的特征

多肽主链上存在三种键(左)。每种键都有各自特定的键长和键角。由于部分双键(即双键在C-O与C-N二个位置间共振，右)的形成，酰胺键不能自由旋转，使酰胺(肽)键的平面得以维持稳定牢固。其他二个键则可自由旋转。(引自 Dickerson与 Geis, 1969。)

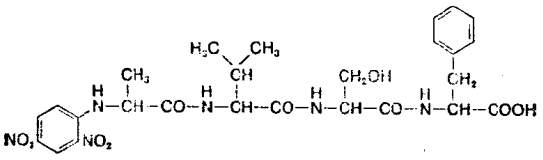
成。）由此可见，蛋白质的一级结构并不能作为阐明蛋白质功能的唯一可靠的依据。

最初被确定氨基酸排列顺序的一些多肽链，例如催产素、抗利尿素、胰岛素等等，同一般酶分子的肽链相比都比较小。当时 Frederick Sanger 为确定这些蛋白质顺序所采用的方法（图 146），以今天自动化分析程序来衡量亦未免过于烦琐和笨拙。但是，作



二硝基氟苯　　游离氨基

↓　Ala—Val—Ser—Phe 组成的四肽



DNP肽

↓　HCl部分水解

不同长度的 DNP 肽

1. DNP Ala
2. DNP—Ala—Val
3. DNP—Ala—Val—Ser
4. DNP—Ala—Val—Ser—Phe

↓

分离每个 DNP 肽段

决定每个肽段的氨基酸组成推导原始肽的顺序

图 146　确定一个四肽顺序的 Sanger 法

2,4-二硝基氟苯(FDNB)与多肽反应时，能结合到多肽氨基端的游离氨基上生成一个二硝基苯(DNP)肽。将 DNP 肽(本例为一个肽)用盐酸温和处理使之不完全水解，就可能产生有四种不同的DNP肽，即相当于原来的DNP肽再加上三种各在不同肽键上断裂生成的DNP肽片段。现在只要测出这种 DNP 肽片段的氨基酸组成，那么原先四肽的顺序就可据此推演出来。这一方法的逻辑推理并不复杂。如附表所示：带有一个氨基酸的 DNP 肽片鉴定为丙氨酸，带有二个氨基酸的DNP 肽片鉴定为丙氨酸和缬氨酸。这就是说，在四肽顺序上丙氨酸的后面一定为缬氨酸，依此类推便可勾划出整个肽链的顺序图。肽链较小时，分析比较简单，如图中所示。肽链过大，程序也相对繁复得多(见下表)。

分离的 DNP 肽片段	氨基酸组成	确认的片段顺序
片段 1	DNP-Ala	DNP-Ala
片段 2	DNP-Ala Val	DNP-Ala-Val
片段 3	DNP-Ala Val Ser	DNP Ala-Val-Ser
片段 4	DNP-Ala Val Ser Phe	DNP-Ala-Val- Ser-Phe

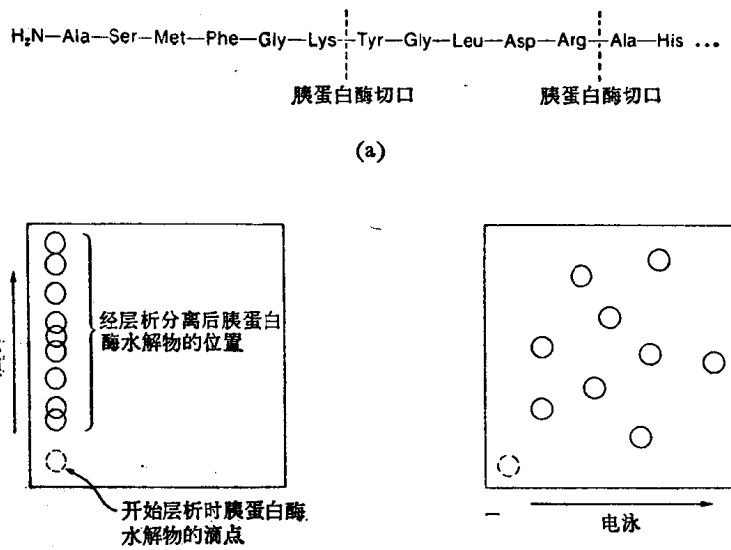
者仰仗于逐个鉴别多肽链肽段上 N-端的氨基酸残基(使其同二硝基苯结合),艰苦卓绝,持之以恒,终于取得成功。

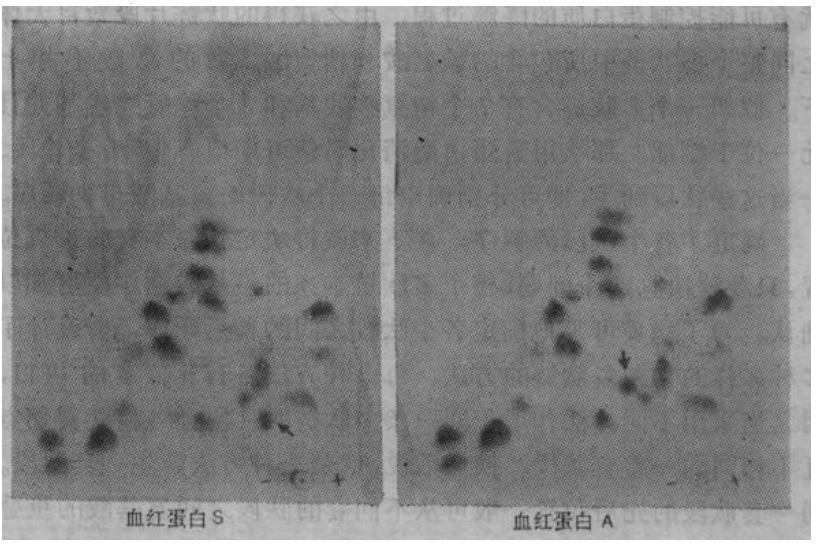
如前所述, Ingram 是第一个阐明遗传突变与蛋白质中氨基酸变换之间存在对应关系的学者。当时,为了研究这一命题, Ingram 必须选择一种便于研究又具有多种突变形态的蛋白质作为研究材料。所以他不可能象 Sanger 那样, 对于所分析的蛋白质大小有较充分的选择余地。Ingram 选用的蛋白质材料是血红蛋白, 突变型血红蛋白则取自贫血症患者的镰形红细胞。血红蛋白是一类较为大型的蛋白质, 这样 Ingram 就面临着一个难题, 这就是首先必须要改进 Sanger 分析蛋白质中氨基酸顺序所使用的方法。

后者为确定血红蛋白氨基酸顺序所制定的策略是: 第一步先把血红蛋白分解为α链和β链 (功能血红蛋白分子由二条α链和二条β链组配而成),然后将多肽链切成片段。这里涉及到使用许多种能成功地降解多肽的蛋白酶,其中有少数为限制酶,能够专一性地在特殊位置上切断多肽链。例如, 胰蛋白酶只在赖氨酸羧基一侧以及精氨酸羧基一侧断开肽链 (图 147), 因此使用胰蛋白酶

就有可能控制蛋白质的降解过程，由之获得的肽链片段数目大体上同整个多肽链中所包含的赖氨酸和精氨酸二者的总数不相上下。假如一个多肽链含有7个精氨酸残基和3个赖氨酸残基并且无一位于端部，那么用胰蛋白酶消化时就可期望产生11个片段。一当这些肽段到手，即可分别确定每一个肽段的氨基酸排列顺序。

确定了各个肽段的顺序，接下来就得决定肽段本身的排列位置，只要解决这两个问题，整个多肽链本来的氨基酸顺序便可加以确认。为了稳妥可靠地确定各个肽段之间的前后顺序，曾设计了多种多样的切断多肽链的方法。每一种方法各有其特定的切口，因之当采用不同方法作用于同一条多肽链时，便会各自获得切点互不相同的一整套肽段。换句话说，就是一种方法产生一套肽段。同一套肽段的先后位置一般可从不同套的肽段之间氨基酸的重叠关系上反映出来（图148）。通常使用的肽断裂方法还有其他一些特异性的蛋白水解酶类和非酶的化学断裂法两大类。糜蛋白酶是其中一种，但它往往不如胰蛋白酶那样稳定可靠，因为糜蛋白酶通





(d)

图 147　从蛋白质降解混合物中将彼此均匀混合在一起的不同肽段进行分离鉴定的方法

(a) Ingram 改进了 Sanger 的方法。使用专一性限制酶在特定的位点切割肽链。例如，胰蛋白酶只切断精氨酸或赖氨酸后面的酰胺键，这样，同一类蛋白质用同一类限制酶切割必定会产生同一类群的肽段。

(b) 将胰蛋白酶消化产生的肽段用纸层析和电泳法分离获得肽谱。做法是在长方形层析纸的一角滴上一滴混合的肽液，然后使用一种溶剂带动肽段沿着层析纸的一边渗透移动。由于每一种肽在溶剂中溶解度不同，因此不同的肽段移动距离(从起滴点计算)也不相同。

(c) 层析纸再置于电场中电泳，该电场以与层析方向成直角方向移动肽段。在电泳中，肽段移动的距离与其所携带的电荷成正比。

(d) 血红蛋白A和血红蛋白S的实验肽谱。将二者比较后发现其中只有一种肽段展现出位置差异(箭头所指处)。

常切断的部位是同芳香族氨基酸邻近的肽键，但有时也会切断亮氨酸和甲硫氨酸后面的肽键，这要看与后二者相邻的氨基酸残基而定。

溴化氰 (CNBr) 在甲硫氨酸残基的羧基侧以非酶方式断裂肽键。由于大多数蛋白质含有比较少量的甲硫氨酸残基，所以用溴化氰产生的肽段一般较大。我们的目标是发展多种具有特异切割点的裂肽方法以便获得二套或多套肽段组(即获得几个肽组)。在

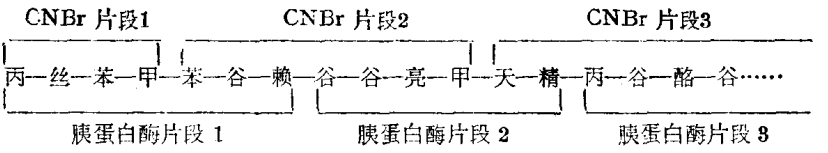


图 148 采用肽段重叠法决定蛋白质顺序

假若采用不同的限制酶或具不同专一性的化学方法消化一个提纯了的多肽，将产生不同的肽组。若在二套这样的肽组中每个肽段均被确定顺序，那么原来的多肽顺序通过比较肽组之间的重叠顺序便可予以确认。

图中一套肽组由胰蛋白酶消化获得，另一套肽组由溴化氰裂解产生，后者切断点是连在甲硫氨酸后面的肽键。如这二套肽组分别提纯并确定其顺序，于是天然蛋白质中肽段的先后次序便可由重叠肽段的关系推导出来。例如，图中CNBr 片段 2 中前面三个氨基酸残基同胰蛋白酶片段 1 的最后三个残基重叠。CNBr片段 2 后面的四个残基同胰蛋白酶片段 2 的前面四个氨基酸重叠。从这种重叠现象可以证明，胰蛋白酶片段 1 和 2 在天然蛋白质中是先后紧密连接在一起的。

不同的肽组之间必定会找到某些氨基酸的顺序相互重叠，因此，根据这些重叠的氨基酸顺序，即可进而确定整个多肽的氨基酸顺序(图 148)。

不同的肽组还可用另一种方式产生：在一种场合下，使多肽链中的赖氨酸残基抗胰蛋白酶，在另一种场合则使精氨酸残基抗胰蛋白酶，这样只要采用同样的胰蛋白酶法就能一举产生三套肽段。这是因为通过特异的化学修饰可分别使赖氨酸残基和精氨酸残基得到对胰蛋白酶的抗性，而带有抗胰蛋白酶赖氨酸残基的肽链只能在精氨酸残基处切断，同样，具有抗胰蛋白酶精氨酸残基的肽链只能在赖氨酸残基处切断。未经化学修饰处理的肽链则是在赖氨酸残基和精氨酸残基二个位置上同时被切断。一般说来，选用什么方法来获得肽组，在某种程度上往往都是根据蛋白质的性质与氨基酸的组成来确定的。

肽段产生之后，普通有二种方法可供确定其排列顺序。一是由 Sanger 设计的用于决定胰岛素顺序的方法。在一定条件下，2,4-二硝基氟苯(FDNB)会同氨基酸的游离氨基起反应。在蛋白质中，较易受影响的游离氨基系位于氨基端的氨基酸残基上。然

后，将用 FDNB 处理过的蛋白质水解，通过层析法分离去除由水解而产生的多种氨基酸。将水解产物中的二硝基苯(DNP)的衍生物取出之后，即可鉴定与之相连的那个氨基酸。最早成功分析胰岛素组成的是 Sanger，作者将胰岛素用 FDNB 处理后分析结果时发现，获得的竟是等克分子量的 DNP-甘氨酸和 DNP-苯丙氨酸。这就说明了胰岛素乃是由二类多肽链组成的蛋白质，一类多肽链的N-端残基是甘氨酸；另一类多肽链的N-端残基是苯丙氨酸。

第二种行之有效的方法是由 P. Edman 研究提出的。根据苯异硫氰酸同肽链的氨基端残基能起反应这一事实（图 149），Ed-

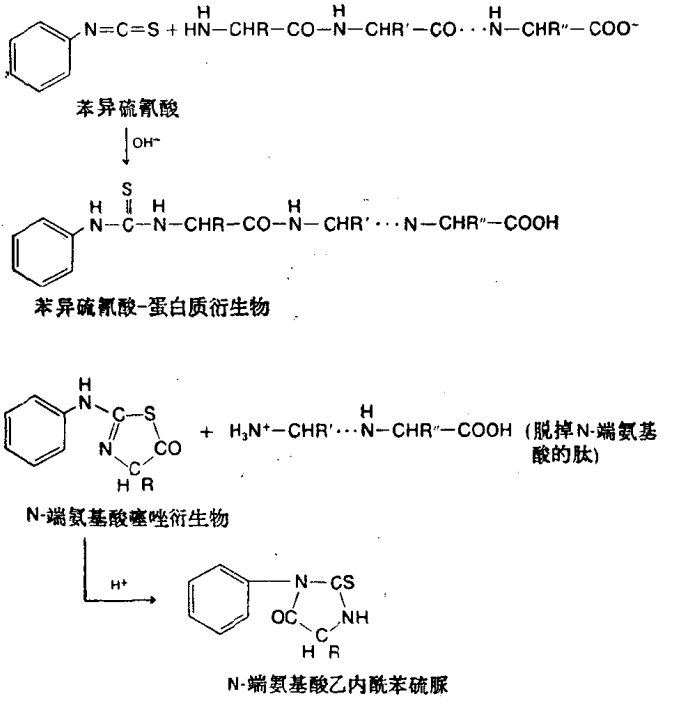


图 149　决定多肽链顺序的 Edman 法

此法是在其他方法基础上的一个改进，因为苯异硫氰酸连接在氨基端氨基酸残基上，从而使得N-端肽键更易酸解。当N-端氨基酸的衍生物解离出来，随即可对该氨基酸进行鉴别。剩余的肽段亦照此办理，这样可使多肽每经一次处理便脱去一个氨基酸，直至全部氨基酸的顺序组成被鉴别确定为止。这一原理已由分析氨基酸顺序的机械化程序(顺序仪)成功采用。

man法也同样能成功地顺次切断 N-端的氨基酸残基。N-端氨基酸残基与苯异硫氰酸作用产生的复合物通过盐酸水解可分离出一个乙内酰苯硫脲衍生物，采用沉淀法即可加以回收。第一个衍生物切除之后，即裸露出第二个氨基酸残基，于是再重复上述的切肽过程。Edman 的顺次降解法与 Sanger 法有所不同，主要差别在于后者在 N-端氨基酸残基移去之后，余下的多肽链仍然保持完整状态。

图 150 列出的是有关两种蛋白质的一级结构，即氨基酸的排

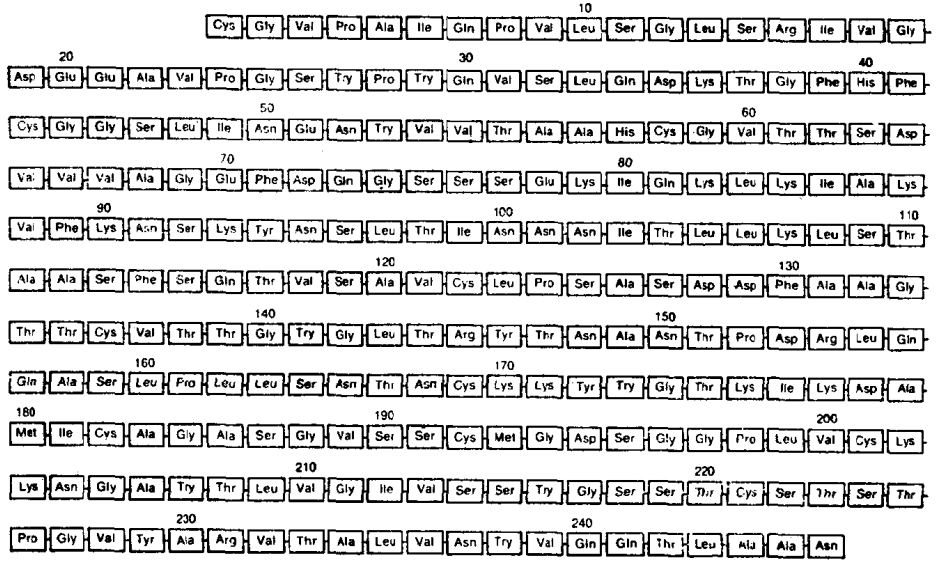


图 150　蛋白质的一级结构

已确定顺序的蛋白质为数颇多。本图所列为胰凝乳蛋白酶原和胰岛素的一级结构。胰凝乳蛋白酶原是一个具有 246 个氨基酸残基的单链(S-S桥未表示)。胰岛素则有二条短链其内部由二硫键相连接。

列顺序。虽然在蛋白质如何从一级结构构成二级结构方面我们已积累了相当丰富的知识，但是迄今为止关于蛋白质一级结构与蛋白质功能之间的关系依然鲜为人知。因此着手考察这些井然有序的复杂结构将会成为人们认识蛋白质催化功能的一把钥匙。

二级结构

任何长链的多聚体都是由不断重复的原子组合单元（即结构单元）连接而成的，此种结构上的周期性会表现出各种对称性模型（对称性是指对长链中任何一点而言，其两边均表现为对称——译注）。因而原子排列上的对称性就规定了长链的基本特征。在一个多聚体的长链分子中，每个亚单位（指组成结构单元的单位——译注）都处于一定的位置。作为大分子整体来说，如将每个重复原子组合中的某个亚单位相互用线条相连，就会看到它们围绕一个共同的轴旋转，所以，亚单元的定位特点在某种程度上亦决定着多聚体大分子的结构。

一个肽链实际上包含一种以上的重复出现的亚单位，故肽链本身算不上真正的聚合物。只有当组成多肽的氨基酸单元彼此极其相似，这样的多肽（如果自然界果真有此种多肽的话）才会具有象真正聚合物（仅由一种单体聚合而成——译注）所固有的某些对称模型的特征。不过在一般的蛋白质多肽链中，也有某种重复的结构单元出现，从而显示出某种程度的对称性。这就是氨基酸残基之间借以相连的酰胺（或肽）键，以及如前所述，在酰胺氢原子和羰基氧原子之间所常产生的氢键（图 151）。从参与组成氢键的酰胺位置来看，肽链的对称方式又可采取螺旋型、折叠型或其他构象（图 152）。肽链的对称性一般能为键的旋转所左右，肽链内有三种键能够旋转（图 145）。每一种键的旋转都受氨基酸侧链位置及其特性所制约，而且原子间除氢键以外的各种弱相互作用对之亦有影响。所有这些力结合在一起从而使整个多肽链的结构和对称性得以奠定，由此而形成的这种结构和对称性模型即称为多肽链的二级结构。

Pauling 和 Robert Corey 首先提出多肽链呈 α-螺旋结构的

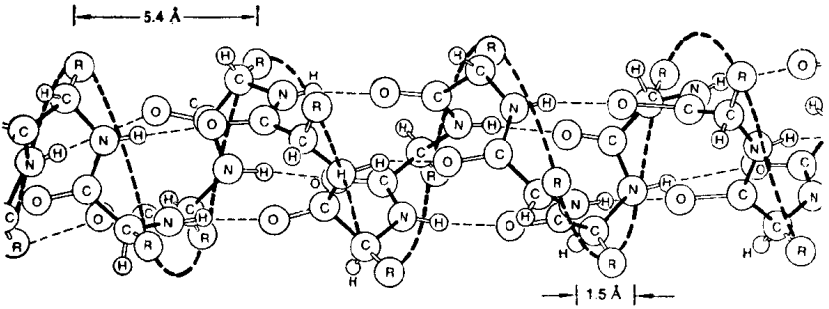
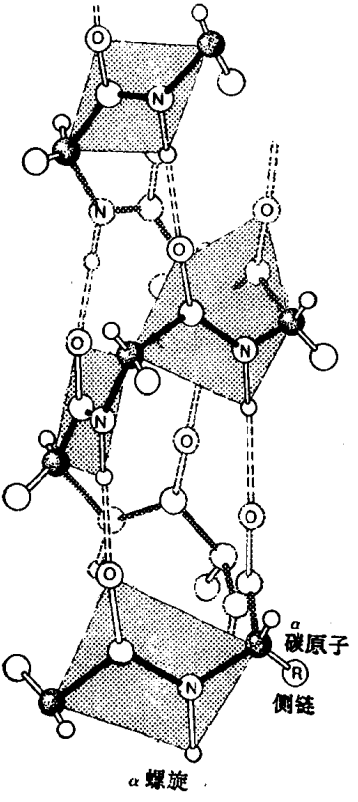
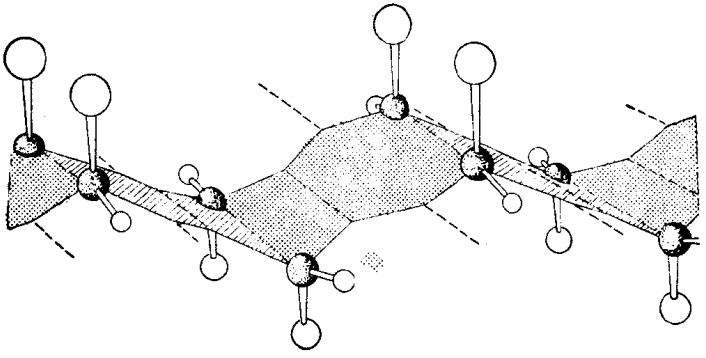


图 151　氢键对二级结构的影响

在已知为 α-螺旋的多肽链构象中，沿螺旋轴方向每隔 1.5Å就有一个氨基酸残基。在每一轮螺旋周期中平均有3.6个氨基酸残基，距离为 5.4Å。衍射图像指示出从一个酰胺氮到与之形成氢键的羰基氧之间一般间隔 3 个残基。强烈的衍射线亦能展现出多肽主链的痕迹。(引自 P. Doty, 1957。)





β 折迭

图 152　α-螺旋结构和 β-折叠结构

在α-螺旋区域里所有主链原子在多肽链内均显示同一取向。氢键总是沿着螺旋轴方向连接着羰基和其后第四个氨基酸残基的氨基。β-构象或折叠结构亦由氢键固定，但形成的氢键同肽的主轴相垂直，α-螺旋的氢键一般与主轴相平行。(引自 Dickerson与Geis, 1969。)

理论。就理论上来说，这一螺旋结构的对称性，既可以为右旋，也可以为左旋(图 152)；但实际上要证明这种螺旋对称性殊非易事。曾经获得使用单一氨基酸作为重复单位合成的人工多肽（真多聚体)，在现有溶剂的条件下此种多聚体分子的二级结构绝大部分为α-螺旋结构(如果不是全部的话)。将天然蛋白质同合成的多聚体肽链加以比较，虽然在前者的某些蛋白质类型中显示为部分 α-螺旋，但对大多数蛋白质而言，这一证据依然带有推论性质。表 19 所列系为每个多肽链中存在的 α-螺旋结构的估计值。

在一个多肽链内，由于 αC-N 键和 αC-C 键的旋转受到限制，可能得到的构象数目亦极为有限。所以只要指明整个肽键上所有的键角便有可能就多肽链的二级结构图象作出比较精细的描绘。这项工作现已由 Pauling 及其小组完成，作者所提出的关于描绘α-螺旋构象所必需的条件如下：

1．每一个羰基氧原子和每一个酰胺氮原子必须包含在氢键结构之中。

2．肽基必须呈一平面，并且键长和键角同简单的酰胺结晶二级结构一致。

表 19　不同蛋白质中 α 螺旋的百分比

蛋　白　质	α 螺　旋　%	测　定　方　法
肌红蛋白	68	圆二色性
	70～80	旋光色散
	65～72	X射线衍射
溶菌酶	29	圆二色性
	30～40	旋光色散
	28～42	X射线衍射
核糖核酸酶	9	圆二色性
	15	旋光色散
	6～18	X射线衍射

3．组成氢键的氢原子必须位于连接参与形成氢键的氧原子和氮原子的联线上。

4．多聚体中每个重复单元或氨基酸残基绕中心轴的旋转方式必须完全相同。

使氨基酸侧链相互连接的氢键，或使酰胺氮原子同羰基氧原子两相连接的氢键，在稳定 α-螺旋结构方面起有重要作用（图152）。可以设想 α-螺旋结构犹如一个圆柱体，而肽基平面则与此圆柱弧面相切。在这样一种构象中，每个肽基的这种定位方式可以使之在主链的上下二个方向上与其后的第四个肽基形成氢键。每隔三个氨基酸残基之间形成的这种氢键一般与圆柱体的中轴接近平行。在整个螺旋长度上，以这种方式间隔排列的氢键起着维持螺旋结构旋向的作用，颇象螺旋形梯子外侧的某种垂直支持物。氢键是原子间一种较弱的相互作用力，由于这一原因，要使这种吸引力切实有效，原子间的距离就势必很短。根据键长和键角的测量，已经推算出每一轮(圈)螺旋周期一般具有 3.6 个氨基酸残基。由此可见，在 α-螺旋构象中螺旋是如此紧密地聚集甚至连水分子亦无容身之地。另外由密度计算，亦同样表明在螺旋轴

内确实没有多余的空间。

第二种常见的二级结构模型称为折叠型，或β构象。这一结构提供了另一类有关对称性构象的有趣的对比（图 152）。折叠型与α-螺旋型结构之间存在的主要差别在于：α-螺旋型中氢键同螺旋的长轴平行，而在折叠型中，氢键则同长轴垂直。蚕丝是一种具有β构象的蛋白质（图 153）。而丝蛋白也有一个重复的氨基酸顺

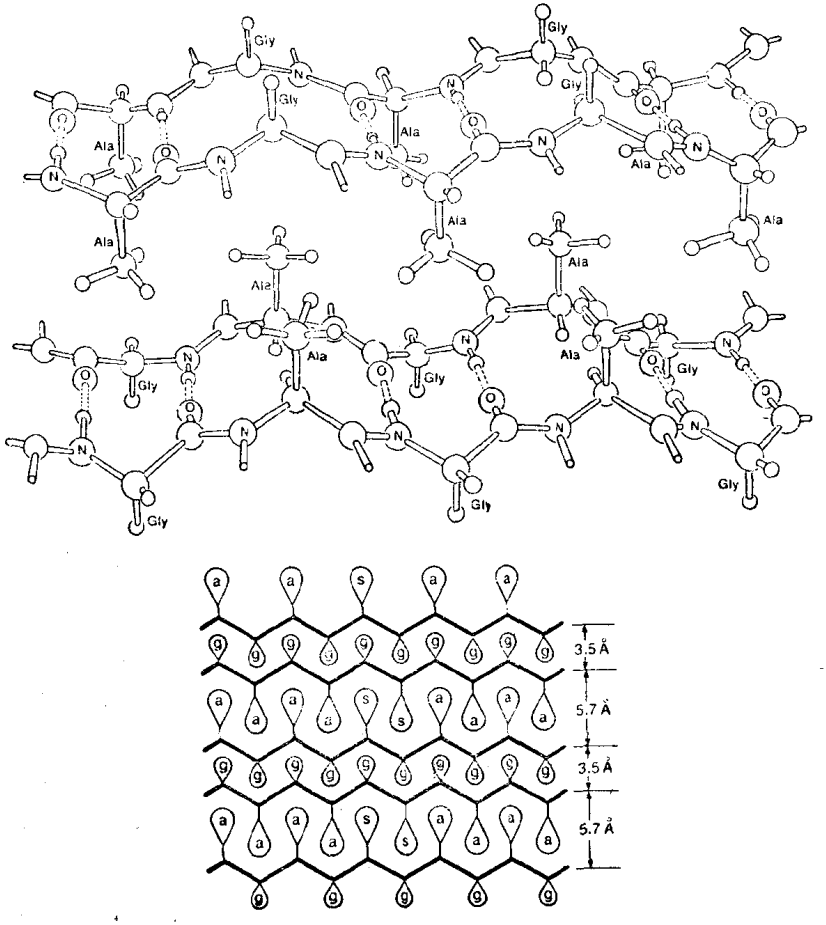


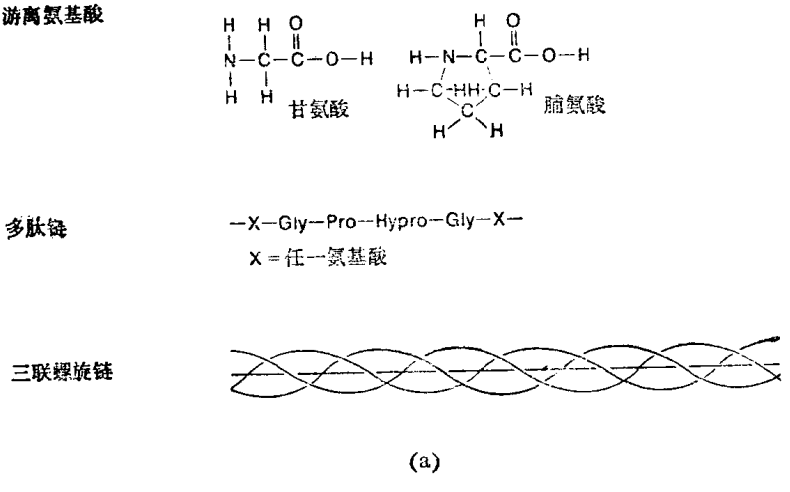
图 153　蚕丝蛋白结构

蚕丝蛋白的二级结构全部为β-折叠型。平行链表示丙氨酸残基对着丙氨酸残基，甘氨酸残基对着甘氨酸残基。平行链之间的间距轮番交替，相应为 5.7 Å 和 3.5Å。（引自 Dickerson 与 Geis, 1969。）

序，重复单位只由六个氨基酸残基(甘-丝-甘-丙-甘-丙)$_n$组成。各折叠层平行排列，结果甘氨酸总是对着甘氨酸，丙氨酸总是对着丙氨酸，丝氨酸总是对着丝氨酸。折叠层之间交替横隔着3.5Å和5.7Å的空间。

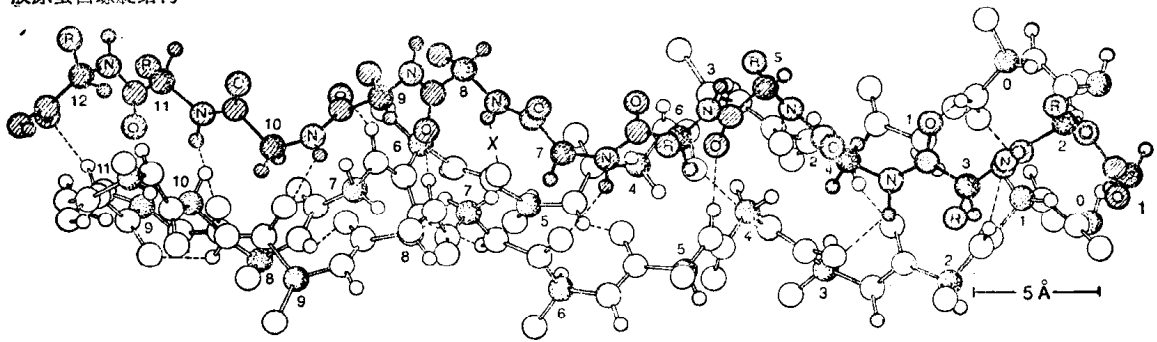
除了α-螺旋和β-折叠这二种主要的二级结构外，还观察到存在有其他类型的二级结构。例如胶原蛋白虽为螺旋结构，但是这种螺旋之所以别具一格，部分乃是因为多肽链内脯氨酸有规则地间隔排列。脯氨酸中特殊的亚氨基侧链使肽链展现出特异性的弯曲。弯曲全部有规则地发生在同一个旋向上，从而造成了另一种类型的螺旋构象。在血红蛋白及溶菌酶中还有一种称为 $^{3}10$ 螺旋结构的构象，它具有完全不同于在 α-螺旋中所见到的键角，而且所有看到的此类结合键所涉及到的多肽链的肽段都非常之短。目前，进行过详尽测定其二级结构的蛋白质尚为数寥寥，因而可以预期也许还有其他类型的对称性模型正待于发现。

通常有二种行之有效的方法可用于鉴别溶液中蛋白质的二级结构类型，这就是旋光色散法 (optical rotatory dispersion) (ORD)和圆二色性法(circular dichroism)(CD)(图 155)。这二种方法都是根据平面偏振光按其通过非对称性介质所显示的特征来



(a)

胶原蛋白螺旋结构



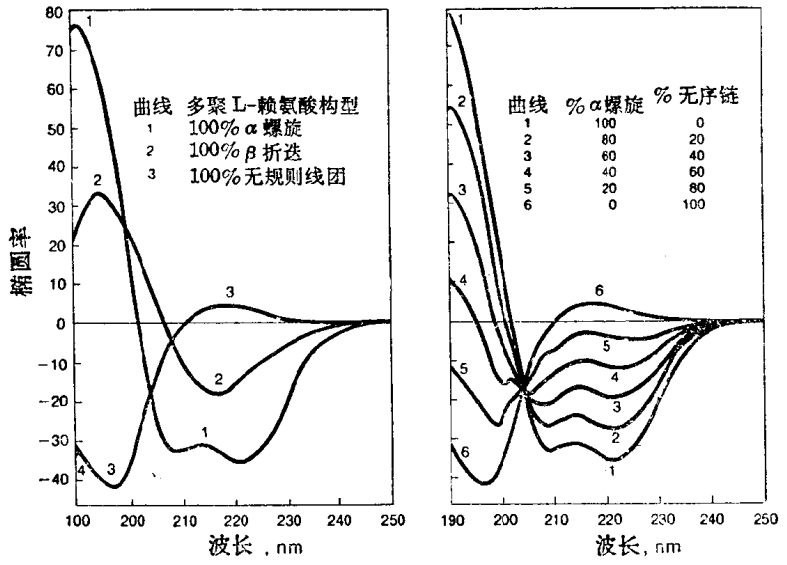
(b)

图 **154** 胶原蛋白结构

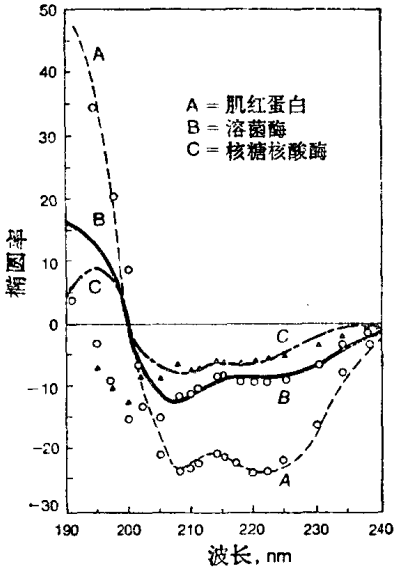
胶原蛋白的分子结构显示为一种由三条左手螺旋的单链相互之间各以右旋方式缠绕进而形成的超级螺旋。(引自 J. Gross, 1961。)

设计的。偏振光一般可分为二个旋转方向的偏振光束，一种为左旋，一种为右旋，每一种均有可测量的不同的物理性质。关于ORD法、CD法和其他各种测定蛋白质特性的分光光度法，对于本章要讨论的问题太过于技术性了，图155中仅就其一般原理略加介绍一二。

关于蛋白质二级结构的问题暂且讨论到此。迄今为止，在制约二级结构的各种因素中，我们已经提到了主链上包括肽键在内的共价键，还提到有助于稳定α-螺旋和β-折叠构象的氢键。除此之外，在组成肽链的原子之间尚有其他一些较弱的力在起作用，即弱相互作用。“弱相互作用”或“弱键”(weak interactions or weak bonds)这一术语常常用来概括所有除共价键之外的原子或分子间的吸引力。对于这些弱相互作用我们还提不出详尽的解释，可以推想它们在蛋白质二级结构中所起的作用总是离不开蛋白质结构和酶功能这些主要特征的。酶的一个最根本性质是它们具有折叠成为一个紧密聚集的独特构象(图171a)的能力，在这种结构中与



(a) (b)



(c)

图 155　采用圆二色性(CD)法测定蛋白质中的二级结构成分

通过蛋白质溶液的圆二色性光谱，能够测量不同波长的透射光所显示的椭圆率变化。椭圆率与透射光中二个圆偏振光组份间的光密度差成正比。由于椭圆率来自于氨基酸残基中不对称碳原子的旋光性，因此在无旋光性的化合物中便看不到此种椭圆率的变化。

(a) 在相应条件下，多聚 L-赖氨酸各以 100%α-螺旋，100%β-折叠或 100%无规线团的形式存在。图示多聚 L-赖氨酸每种构象的 CD 光谱曲线。

(b) 三种构象的多聚 L-赖氨酸按任意比例相互组合，并测定每种组合中各构象的百分比。图示为 0%的β-折迭与不同百分比的α-螺旋和无规线团组合的光谱曲线。

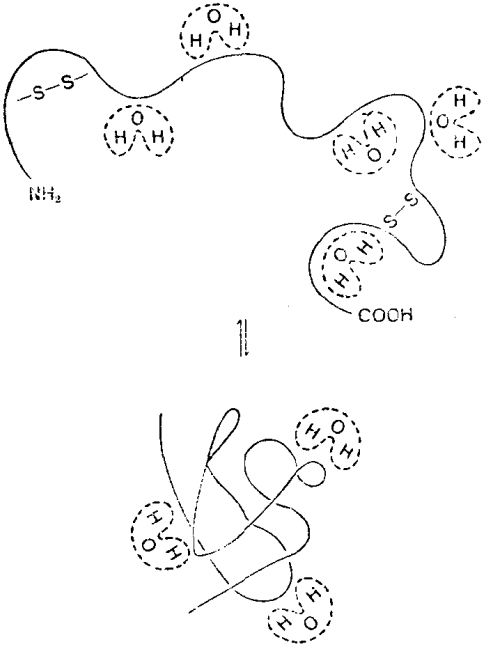
(c) 组合光谱曲线的实际应用。图中数据(列点)是对肌红蛋白、溶菌酶、核糖核酸酶这三种天然蛋白质实际观察到的 CD 光谱。通过列点描成的曲线同(a)中组合曲线颇相符合。这里与肌红蛋白(曲线A)相当的一条曲线经计算其组分为 68%的α螺旋，5%β折叠和 27%无规线团。与溶菌酶(曲线B)相当的曲线组合为 29%的 α-螺旋，11%的β折叠和 60%无规线团。与核糖核酸酶(曲线 C)相当的曲线组合为 9%的α螺旋，33%β 折叠和 58%无规线团(参见表 19)。(引自 Greenfield 与 Fasman, 1969。)

反应直接有关的氨基酸侧链将构成具有高度特异性的能和底物起反应的结合部位。在水溶液中球蛋白大概也是处于某种“折叠-舒展”不断反复的可逆状态，即来回经历着从杂乱无章、线形伸展、未经折叠的多肽链到通过折迭而叠成紧密有序的、多重折叠

的、具有酶活力而接近球状的构象的过程。所有这一切恐怕都与此种弱相互作用不无关联。总之，作为蛋白质二级结构的基本要素，各类蛋白质都基本相同，但是它们在分布和折叠形成三级结构的方式方面，每一种蛋白质却又各有千秋(参见下节)。

从原子之间的相互作用方式来看，以形成牢固的共价键为一个极端，相互的排斥性为另一个极端，表现形式多种多样，作用范围也十分宽广。在多肽链中，主链上的共价键决定着分子的线型结构，而弱的相互作用，特别是氨基酸侧链间的相互作用则对多肽分子的形态有相当的影响(图 156)。

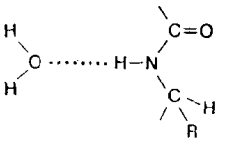
在原子的弱相互作用方面了解得最为清楚的应推氢键，它们往往在带有若干正电的氢原子和带有若干负电的其他原子之间出现，此类带负电的原子在分子内几乎总是以共价键的形式同其他的原子相结合。通常与氨基酸侧链中的氮原子或氧原子结合的氢



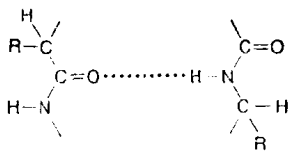
(a)

未折迭形态中蛋白质-溶剂相互作用

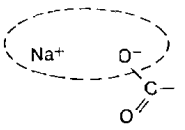
折迭形态中蛋白质-蛋白质相互作用



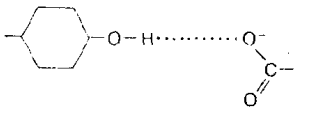
水-蛋白质氢键



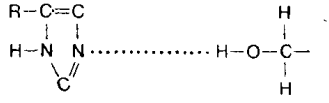
肽内氢键



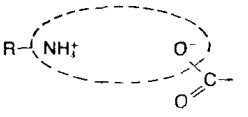
溶剂-蛋白质离子键



羰基与酪氨酸羟基间氢键



组氨酸和丝氨酸侧链间氢键



(b)　　蛋白质-蛋白质离子键

图 156　非共价键相互作用

溶液中的蛋白质分子可以从无规则的自由伸展的状态转变成搭配有序的折叠构象。在这一过程中蛋白质分子内部的相互作用即取代了蛋白质-溶剂的相互作用。天然状态(即折叠型)的蛋白质为了维持其有序结构所必需的能量，会从非折叠多肽链(即无规则的自由伸展的多肽链——译注)同水之间相互作用的去稳定化所释放的能量得到补偿。在从非折叠状态过渡到天然蛋白质状态时，不论从范德华力、氢键，还是从离子键都没有获得净增的稳定化能量，这是因为这些相互作用力在非折叠和折叠二种状态中都接近相等。天然状态蛋白质其结构极为紧密(象结晶中的氨基酸那样靠近)，因而蛋白质-溶剂相互作用的范德华力在二种状态中差异不大，折叠型蛋白质内部所有原子都"各得其位"，极其致密(图 171a)。此外由于原子之间密切聚集，绝大多数内部极性基团亦会形成蛋白质分子内的氢键，而分子外部的极性基团因依然同水分子间维持氢键，故在非折叠状态和折叠状态中氢键的总数目亦不相上下。从蛋白质-溶剂间的离子键来看，在二种状态中也都差不多相等，带电基团一般都位于球状蛋白质外部，它们与溶剂中相反电荷的离子保持着相互作用。

一般认为，多肽链从未折叠的无规则的状态转变成高度有序的蛋白质分子所引起的熵值(熵能)减少的补偿过程，可能与溶剂水分子从紧密聚集的蛋白质基质内部排出有关。这一过程的物理学机理还不了解，目前往往用疏水性相互作用的理论作解释：即围绕着未折叠多肽链疏水性基团的水膜，在氨基酸的疏水(非极性)侧链折叠而卷入蛋白质分子内部时遭到破坏，从而及时提供了折叠构象所需要的稳定化能量。

原子可能是作为供体同另一个氨基酸上带负电荷的原子起作用。受体分子则多半是氮原子和氧原子。

溶解蛋白质的溶剂对蛋白质分子中弱相互作用的特性常常表现出深刻的影响，事实上极性溶剂常常以氢键形式同氨基酸的侧链相互作用。另外，生理溶液的主要成分——水就具有作为氢键供体以及在分子的负电性区域(氧原子)作为氢键受体的能力。

除氢键之外，还有静电结合力、疏水性相互作用力和范德华(van der Waals)力等等。静电结合发生于携带相反净电荷的离子基团之间。在蛋白质中，静电结合则可以出现在侧链的$-NH_3^+$和$-COO^-$基团之间，例如在赖氨酸和谷氨酸中就有这种结合形式存在。但是，关于静电结合仅仅只是对 α-螺旋结构的形成有作用呢，还是对某些其他结构同样也有作用，目前还不得而知。

丙氨酸、缬氨酸、异亮氨酸、亮氨酸、色氨酸、酪氨酸和苯丙氨酸都是非极性氨基酸，因而在诸如水这样的极性溶剂中不易使之溶解。它们象其他非极性分子一样也是疏水性的，并且彼此间的亲和力远远高于同水和其他极性溶剂的亲和力。当蛋白质折叠成球状的三维结构时，非极性侧链由于疏水性作用向分子内部相互聚集，其他极性氨基酸侧链向外伸展到达球形分子的表面，甚至突入分子周围的溶剂。非极性侧链相互聚集从而把溶剂分子自蛋白质内排挤出去的特性称为疏水结合。关于疏水结合的作用本质目前尚存在一些争论，但一般认为疏水性相互作用在蛋白质结构中所起的作用是相当重要的。

要精确说明疏水性作用力、静电键和氢键在蛋白质二级结构中所起的作用，虽有不少困难，但是由于这些作用力各有其特异性，所以即使是在蛋白质这样复杂的大分子中，它们所处的位置和作用强度还是能够测定的。相比之下，范德华力就显示不出特异性。范德华力的作用强度同原子间的距离大小有关，而且不局限于分子的某些特定区域，在极性分子、非极性分子以及携带电荷的分子中都有存在。当原子之间相互极为靠近时，此时相邻原子外层电子轨道中的电子云发生重叠，排斥力将会超过吸引力，因此，

排斥力和吸引力作用方向相反，总是会在每对原子间某一特定的距离上达到平衡。由于这二种力的非特异性，加上强度又相对较弱，所以一般颇难判断它们各自对维持蛋白质构象所起的作用。

当进一步考察蛋白质分子的高级结构时，必须充分考虑其相互组合会关系到氨基酸排列顺序和多肽链二级结构形成的全部因素。此外亦不应忽视那些能决定和(或)维持蛋白质二级结构的原子间各种强的或弱的吸引力，这是因为蛋白质的三级结构乃是所有这些因素共同作用的结果。

三级结构

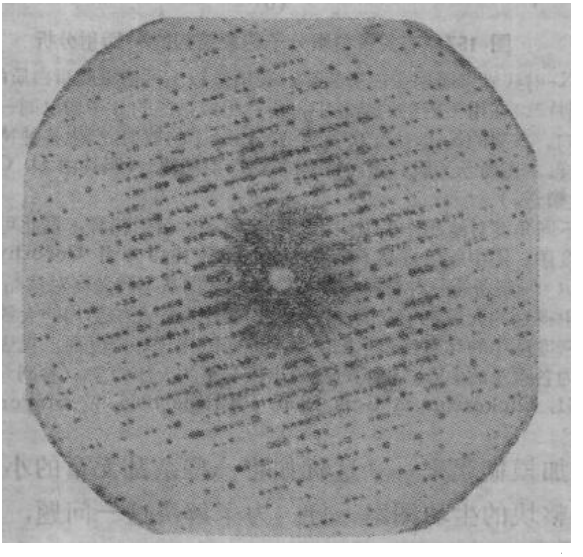
人们早先知道，大多数具有催化功能的蛋白质大体近于球形。后来，由于发明了超速离心机，能够在引力场中将这种具催化功能的蛋白质与其变性的复本（即将原来的或天然的蛋白质加以变性使之成为充分伸展的多肽链)在沉降系数方面进行对比，这一点才最终得到证实。一个充分伸展的多肽链与折迭起来的大分子蛋白质相比，要远为不对称得多、结构紧密犹如圆球状的蛋白质分子具有较高的对称性，因而在引力场中沉降速度更快。虽然已发现许多球状蛋白质的存在以及这种球状蛋白质可以产生变性的事实，但这些并不足以揭示多肽链折叠过程的本质和蛋白质分子为维持三级结构所必需的那些原子间相互作用的方式。另外，蛋白质的结构究竟是固定不动，还是弹性多变？如果确实多变，它又能变到什么程度？通过何种机制来完成这种变化？所有这些问题都难以从中找到解答。

人们关于蛋白质三级结构的认识主要来自对蛋白质结晶进行的 X-衍射研究(图 157)。那么，蛋白质结晶的性质能否真实反映出"活性蛋白质"本身固有的某些性质呢？自然，这个问题还不能最后定论。尽管如此，我们从蛋白质三级结构的 X-衍射所得到的资料毕竟还是为研究蛋白质的功能找到了大量的线索和论据。就这点来说，X-衍射研究对于阐明蛋白质功能的重要性完全可以与 DNA 双螺旋结构对于认识 DNA 功能的重要性并驾齐驱。

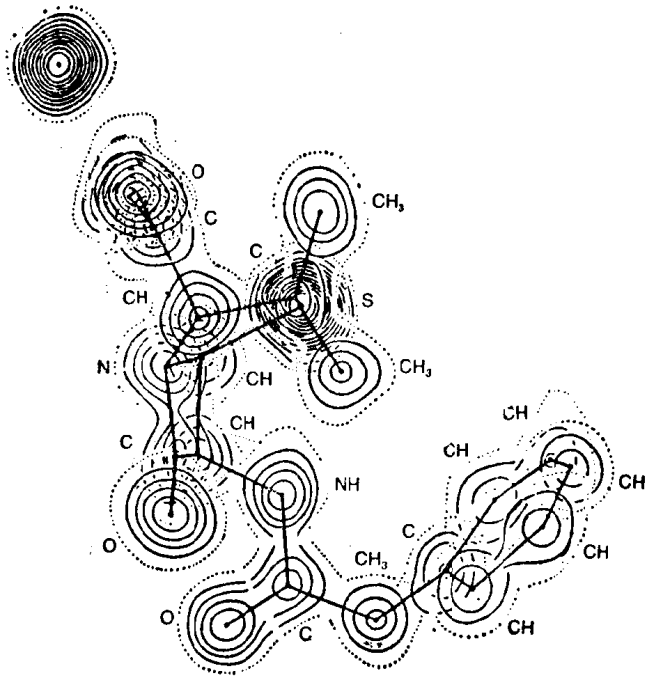
X-衍射的分辨力约为 1Å 左右，小于大多数以共价键相结合

的原子间距。因此，要分辨蛋白质分子中各个原子的排列位置，并进而说明所有原子之间的空间关系，在理论上是完全可行的。在测定样品是小分子化合物的情况下，实验资料基本上可以满足理论推算的需要。但如测定样品是大分子化合物，则难以做到这一点。因为要分辨大分子中心的原子如何排列，没有大量的电子密度图是难以设想的。一般位于 X-衍射图中心的衍射斑点对应于较长的原子间隔，而靠近衍射图边缘的衍射斑点则对应于较短的原子间隔。显然，只有对应于原子间最短间隔的衍射斑点才有可能对原子之间的空间方位发挥最佳的分辨力。例如，为了能够分辨相隔距离达到 2Å的原子在空间的排列位置，对于肌红蛋白这样大小的分子便需要"拍摄"多达 20,000 张"衍射花样"(照片)(系指从不同角度对同一结晶进行 X-衍射——译注)。然后，再把这些照片按三维空间的角度统一起来，这时整个分子中各个原子之间的相互关系才能得以有效的展现。

尽管大量衍射花样的处理已因采用计算机而有所简化，但是随着分析部位从衍射花样的边缘向中心移动，原子的排列方向也



(a)



(b)

图 157　有关蛋白质分子构象结构的X-衍射分析

(a) 用X-射线光束照射蛋白质结晶，在摄像板上可拍摄到蛋白质的X-衍射"照片"。在用于衍射的结晶中，所有组成结晶的分子均以同一方式排列取向。入射的X-射线光束受到结晶中原子的散射，产生某种特有的衍射花样。本图所示为肌红蛋白结晶的X-衍射图片。(照片由 D. C. Phillips 赠送。)

(b) 从不同角度对结晶作X射线衍射可获得多种衍射花样，据此可绘制电子密度图。图中所用样品不是蛋白质，而是1949年由 Dorothy Crowfoot 所制备的一种青霉素钠盐结晶。根据从不同角度对结晶所作的X衍射得到的大量电子密度图，可以勾画出分子三维结构的大致轮廓。电子密度图中的环形线表示围绕各个原子的电子等密度面，直线所连接的点为各原子的中心。由此可展现出该分子的大体构象。(参阅 Kendrew 1961, Dickerson 与 Geis 1969, Phillips 1966, 及 Stryer 1968。)

就变得愈加模糊起来——这犹如将一些杂乱无章的小块拼板玩具去拼凑成整块的生动图案一样。为了解决这一问题，必须在分子的模糊区域先建立一系列能够由X-衍射加以辨认的"测标"，以此作为定位分析的基点。这一方法已由 J. C. Kendrew 和 Max

Perutz 设计成功，其原理是，采用重(金属)原子作为标记物，在蛋白质分子的结构中引入重原子从而得到一系列不同的标定点（图158)。通过这种同晶置换的方法，可以分析比较衍生物分子结晶（含有重原子的分子）和原有天然分子结晶在衍射花样之间的区别，从而能将各个重原子准确定位并确切无误地标出与之对应的相角。由于蛋白质分子中带有许多不同类型的"测标"重原子，即可据以定出各种相角。当蛋白质分子中重原子数目多达一定程度时，所有的相角都能够显露无遗。

采用同晶置换方法已在血红蛋白分子结构中成功地引入了六种不同的重原子。通过X-衍射，得到了所有用于建立三维结构所必需的相角。目前，根据从血红蛋白分子以及少数其他酶分子所

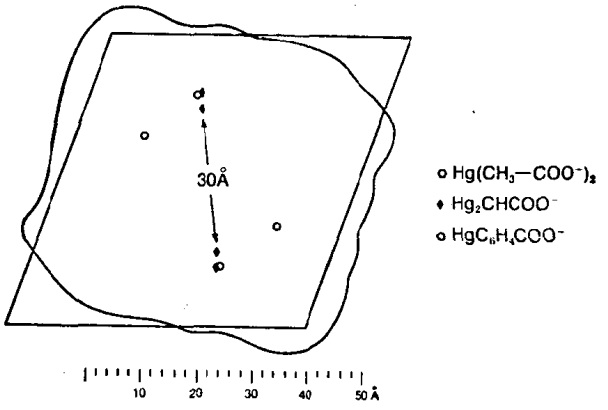


图 158　重原子标记

在蛋白质分子结构中引入重原子，通过X-射线衍射即可确定重原子在分子中的排列位置。由于重原子在电子密度图上的图影与它们特征性的相角相对应，而相角又是特定发射源发出的X-射线光程差的度量。(每个波都可作为一个矢量，矢量长度与振幅成比例，它的方向由相角决定。)因此引入蛋白质分子中不同位置的重原子便会产生不同的相角，如果能制得含有足够种类重原子的蛋白质，那么所有的相角将能有效地显示出来。最早用来验证此法确系行之有效的复合物是高铁血红蛋白加入2M 苯甲酸汞($HgC_6H_4COO^-$) 以及高铁血红蛋白加入2M醋酸汞$(HgCH_3COO^-)_2$混合后所得的产物。图示高铁血红蛋白分子中重原子的相对位置，重原子聚集在二对巯基周围。当重原子与巯基结合时，它们在血红蛋白中的位置距离30Å。图中曲线是血红蛋白分子的大致轮廓。(引自 Cullis, Dintzis 与 Perutz, 1961。)

获得的三维结构资料，已能为蛋白质分子的三级结构模型勾划出一个总的概貌。

一般说来，酶分子只在水溶性的环境中才表现出催化功能。因此人们由此设想，酶催化反应的化学特性应当依从于溶剂的化学性质。但是，从衍射分析提供的三级结构模型来看，并不如此。同水溶液相比，似乎酶分子为其底物提供了一个极性更少的环境。这是因为非极性氨基酸相互作用形成了"疏水性"活力中心，而且这种疏水性结合方式还起着维持蛋白质分子三级结构的功效。

抹香鲸的肌红蛋白分子犹如一个盒子，其大小为 44×44×25Å。盒子里含有一个亚铁血红素基团(图 159)，整个分子结构基

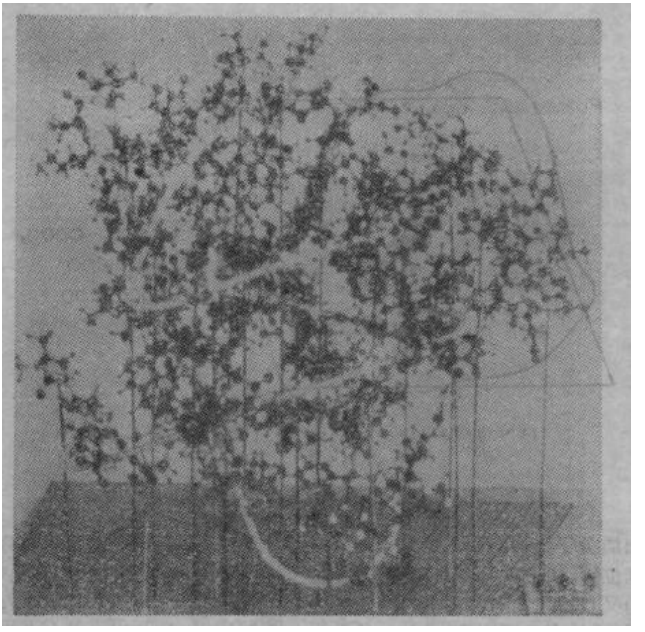


图 159　肌红蛋白的三维结构

这里图示的肌红蛋白立体结构模型真实地再现了肌红蛋白分子内部各个原子的排列位置及其相互关系。分子的结构规则致密，一般情况下总有四个水分子包含在分子内部。几乎所有的极性氨基酸残基都位于分子的表面并同溶剂接触，而溶剂水分子则同暴露在分子表面的极性基团结合。非极性基团大都位于分子内部。读者注意该分子中还含有一个亚铁血红素基团。(本照片由 John Kendrew赠送。)

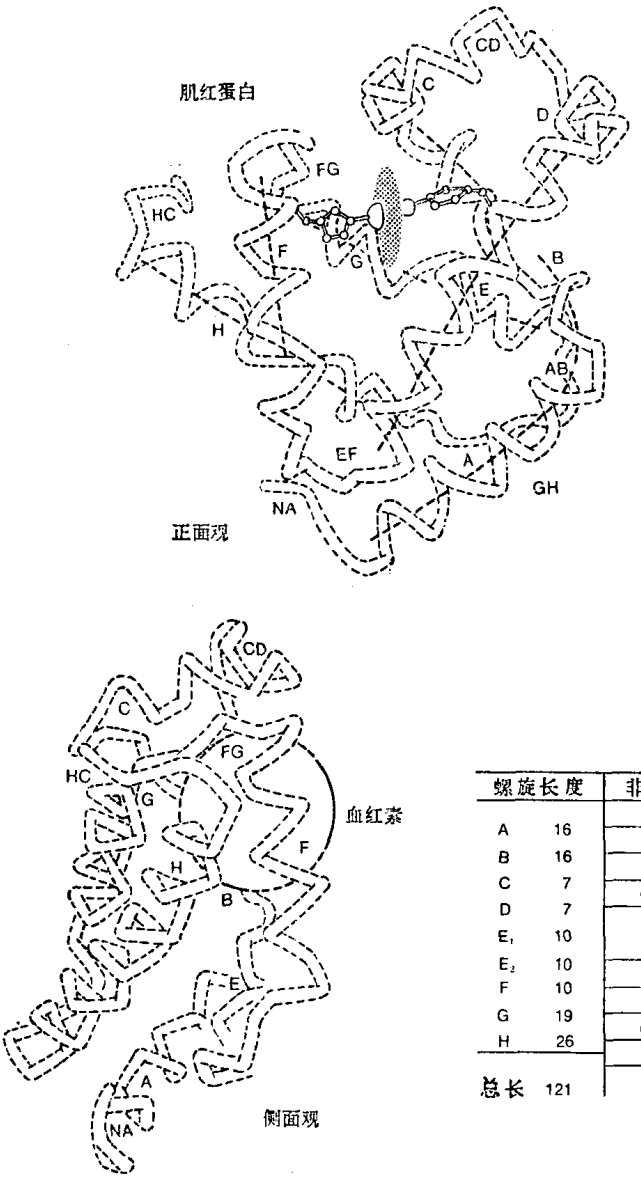
本上是由一个具有 8 段α-螺旋的多肽链折叠聚合而成，这 8 段螺旋的长度分别在 7 至 26 个氨基酸残基之间。肌红蛋白分子共有 153 个氨基酸残基，其中有 120 个在 α 螺旋区。α 螺旋之间彼此反折聚迭，有些肽链折叠区全部为 α 螺旋，而另一些区段则全为非螺旋，其中有二段非螺旋区分别同各自二端的 α 螺旋相连。脯氨酸残基由于其结构上的特殊性正好位于肌红蛋白分子的四个转角上(由于脯氨酸残基含有吡咯环，会导致肽链转折——译注)。

在进化过程中，由于自然选择的结果，使血红素基团的携氧能力日趋完善。8 个α螺旋区域的巧妙构件为血红素基团与肌红蛋白分子的结合提供了一个极好的场所 (图 160)。除血红素上二个丙酸基外，该基团的其余部分正好嵌在肌红蛋白分子的空穴里。这一空穴看来象是一种精巧的保护装置，它还为血红素基团与亲脂的氧分子结合创造了良好的环境。血红素基团所在空穴四周围绕着疏水性氨基酸侧链，或者说由 α 螺旋肽链上的疏水性氨基酸侧链构成了空穴的内壁。而亲水性氨基酸侧链都毫无例外地位于分子的外表面(图 159)。

肌红蛋白结晶的 X-射线衍射图还揭示了血红素基团的几何形状以及它同蛋白质分子的结合方式。铁原子被一个六角八面体所包围，它们提供了一对电子给铁原子。六角之中有四个为氮原子，系由组氨酸供给。这四个氮原子只提供共价键，血红素基团借此得以固定在蛋白质结构当中。这些共价键可优先断裂，使得在不改变蛋白质构象的情况下可以脱去和再接上血红素基团*。整个血红素分子的移去还需要断裂血红素上的丙酸侧链和组氨酸分子杂环上的氮原子以及精氨酸残基上的一个氮原子之间形成的氢键——这是较容易做到的化学"外科手术"。

---

* 一般认为，铁原子位于血红素基团的正中，并同卟啉环上的四个氮原子 (位于四个吡咯环上)以配位键相连而共处于一个平面。在平面的二侧各有一个组氨酸由咪唑侧链上的氮原子以配位键同铁原子结合。因而铁原子实际上由六个氮原子所包围，形如一个六角八面体。通过卟啉环上非极性基与空穴的非极性氨基酸侧链相互作用(非共价键) 以及通过平面二侧组氨酸提供给铁原子的配位键，使血红素基团借以固定在肌红蛋白分子结构中。——译注



螺旋长度		非螺旋长度	
		NA	2
A	16	AB	1
B	16		
C	7	CD	8
D	7		
E_1	10		
E_2	10		
F	10	EF	8
G	19	FG	4
H	26	GH	5
		HC	4
总长	121		32

图 169 肌红蛋白肽链结构的正面观和侧面观

肌红蛋白含有 8 个明显的包围着血红素基团的 α 螺旋区。血红素的两侧都和组氨酸残基的侧链相连接(氧分子结合在血红素的右侧，如正面图所示)。α螺旋区和非螺旋区在插图中已作标明。(引自 Dickerson 与 Geis, 1969。)

肌红蛋白和血红蛋白在一级结构以及生化功能方面极为相似，在三级结构水平上二者也有许多共同点。不同之处是，肌红蛋白一般仅由一条多肽链（单体）组成，血红蛋白则由四条多肽链（四体，即二条 α 链和二条 β 链）组成。图 161 所示为血红蛋白的 α 链和 β 链的结构，它们都具有相似的折叠方式，并且与肌红蛋白分子的构象不相上下。血红蛋白的 α 链同肌红蛋白相比缺少一段含有六个氨基酸残基的 α-螺旋，另外在C端还缺失了六个氨基酸残基（连缺少的螺旋在内共少 12 个残基）。血红蛋白的 β 链也比肌红蛋白少七个氨基酸残基，其中六个也是在C端丢失，还有一个则在N端丢失（因为在N端增加了一个氨基酸残基，而在N端的另一位置又失去了二个氨基酸残基，总计减少一个残基，）。α 链也有与此相同的情形，即在N端某处增加一个残基，在另一处又丢失二个残基，总计亦减少一个残基（连同上述所缺失的 12 个氨基酸残基在内，α 链比肌红蛋白共减少 13 个氨基酸残基。肌红蛋白的氨基酸残基数目为 153，照此计算，α 链氨基酸残基数目应为 140，这与一般记载的 α 链氨基酸数目为 141 的说法不符——译注）。

前面开始讨论X-衍射时，我们就提到关于结晶蛋白质的三维空间构象是否等同于这个工作酶原来的三级结构问题。人们从蛋白质结晶的X衍射研究中，曾经收集了有关这方面的资料，包括将携带底物的酶同不携带底物的酶进行对比等等。结果表明，在血红蛋白的四个单体中只有 α 链和 β 链之间保持接触，而在二条 α 链之间，或二条 β 链之间则似乎并无联系。将上述观察再对照如下二点重要事实综合分析：携氧血红蛋白和去氧血红蛋白之间存在有一定结构差异；α 链和 β 链所构成的半分子在功能表现上类似于一个刚性单位。这样，当从携氧血红蛋白分子与去氧血红蛋白分子的结晶衍射勾划出来的分子构象来推论体内血红蛋白时，便自然会联想到，酶的“工作”（即酶的催化反应，其中主要包括酶分子同底物以及产物之间所发生的连续的结合和释放反应）方式必然会涉及到酶分子内部的运动。按照上面的分析，这种运动势必产生于血红蛋白分子的 α 链和 β 链分子半体之间前后的“滑动”，

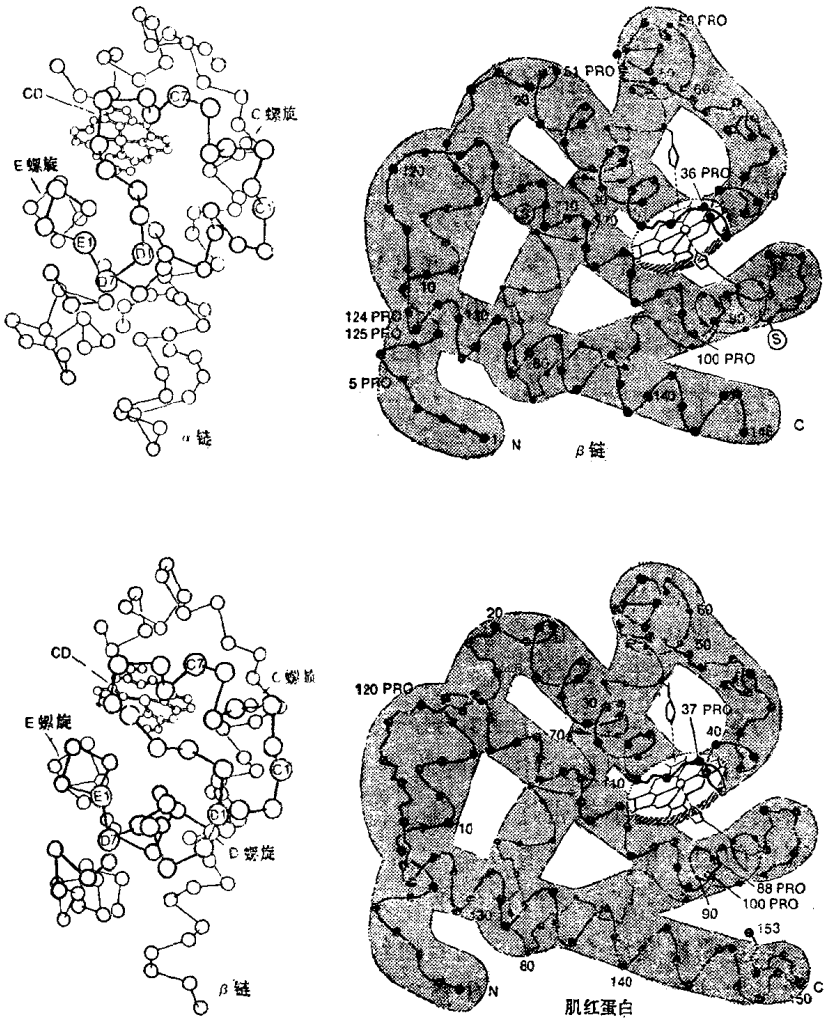


图 161　肌红蛋白和血红蛋白的 α 链和 β 链

虽然肌红蛋白和血红蛋白的α链和β链的一级结构有差别，但在它们的三级结构之间存在有显著的相似性。肌红蛋白取自抹香鲸肌肉，血红蛋白取自马血。海豹的肌红蛋白和人的以及牛的血红蛋白结构也大体类似，似乎都有特殊的折叠方式使血红素基团得以固定并可携带氧分子。（引自 Dickerson 与 Geis, 1969。）

结果使二个 α 链相互靠拢 1 Å 左右，在另一处使二个 β 链分开约 6.5Å左右。采用其他蛋白质分子也进行了同样的研究，情况完全和血红蛋白分子类似。在所研究的这些酶中还有一种溶菌酶，通过 X-衍射揭示出，这个酶分子的内部有一区段在催化反应过程中发生了很有意义的位置变化。这在"酶的作用机制"一节中，还要详细讨论。

蛋白质结晶衍射分析向人们清楚地展示了处在折叠状态中的多肽主链每一区段的走向。除此之外，尚可据以推知每个原子在蛋白质分子中的排列位置，并可确定二硫键所在的部位，决定哪些氨基酸位于分子内部，哪些氨基酸位于分子外部，以及推测由于蛋白质同底物分子的结合引起各部分相对位置的变动等等。至于最初的多肽链究竟是如何折叠的问题，从衍射分析仍然无从得知。氨基酸残基的排列顺序包含着控制折叠过程所必需的全部信息，这一观点得到下述观察的有力支持。实验证明，许多蛋白质分子可以自发地和可逆地折迭成具有它们各自生物活性的构象，这种构象由分子内部的非共价键和二硫键所维持。不妨设想一下，如果将一根 10 英尺的带子塞入一个火柴盒，便会看到带子在火柴盒中的各种可能安置形状将会是一个天文数字。可是类似于带子的多肽链当盘成一个如同小形的盒子时，都总是按照一定的信息有规则地折叠成相同的或者极为相近的构象。

众所周知，折叠过程总是表现出高度协调，即随着多肽链的某一部分开始折叠，随后剩余的部分就能更容易更迅速地完成折叠过程（图 162）。如同所有协调过程一样，第一个过程的开始加速了其后的折迭过程。因此，人们不无根据地设想：或许多肽链上的某一核心部位首先启动并导致了折迭过程的进行。（核心部位被认为是由一定区域中的氨基酸自发形成的结构中心，至于哪些肽链片段处于该区域之中，一般并不明了。）

多肽链进行折叠时，氨基酸侧链各自以疏水性侧链背水进入折叠分子的内部，并且相互聚合。这是最初引起蛋白质多肽链产生折叠的一个策动力。氨基酸疏水性侧链相互聚集，在结构上显

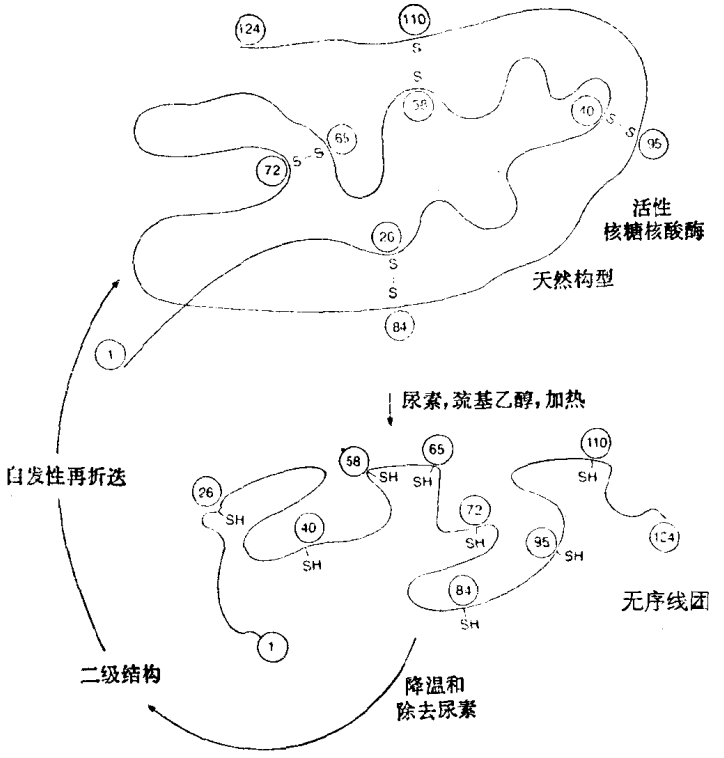


图 162　核糖核酸酶的变性及其折叠构象的复原

核糖核酸酶在外界因素作用下变性，当外界因素消除时又可复性，生成一个具有四个二硫键的折叠多肽链，并且仍然保持着完整的酶功能，在温和的氧化条件下，复性是缓慢的，但却是准确无误的，即二硫键总是在 26–84，40–95，58–110 和 65–72 这些半胱氨酸对之间形成。尚若氧化剧烈的话，则以另外的组合形成二硫键，结果得到无酶功能的蛋白质。实验证明，因二硫键全部断裂所产生的变性核糖核酸酶可以在温和条件下回复到它原来的天然折叠状态，这一事实提供了第一个关于蛋白质一级结构决定三级结构的确切的证据。

得极为紧密，这种结合方式给蛋白质分子的空间构象带来了种种限制，从而使多肽链最终形成一般常见的、能量上许可的即稳定的天然折叠构象的蛋白质分子。

只要蛋白质一级结构和三级结构的关系得以查明，成功合成具有催化和控制功能的新型蛋白质将是一个合乎逻辑的结果。例如，胰岛素由中国率先合成之后，又由德国的小组和美国的小组相

继合成成功。最近，Robert Merrifield 采用逐级固相技术也一举合成了完整的核糖核酸酶(由 124 个氨基酸残基组成)。这些天然蛋白质的人工合成，说明在生化技术上已有相当出色的进展，但这些成就还没有能从根本上阐明蛋白质分子的一级结构和三级结构的内在联系，这仍是今后一个有待努力解决的重大课题。

在生命体系中，蛋白质合成并不是一步完成，一般需要经过一段加工过程。这种加工步骤看来可能比以往所想象的更为普遍常见。糜蛋白酶原是一种无活性的蛋白质，刚合成时只是一个简单的线型多肽链。在该酶原折叠成三级结构之后，通过加工切除了两个二肽使整个分子形成了三条由二硫键连接的多肽链从而得以活化。现在已知，胰岛素也经过相似的加工过程，即胰岛素被合成为一个简单的多肽链之后，又被切去一段含有 30 到 33 个氨基酸残基(依生物种属而异)的片段。详尽了解蛋白质加工过程的某些机制将会加深人们对蛋白质三级结构原理的深入理解。

四级结构

有些工作酶象肌红蛋白那样，由单个多肽链折迭就具有活性；另一些工作酶，则象血红蛋白那样，需要多个折叠的多肽链聚合才具有活性。一般来说前一类酶在数量上要比后一类酶稀少得多。单个多肽链的蛋白质称单体；具有二个或更多单体的蛋白质称多聚体。由二个或多个相同单体组成的多聚体称同源多聚体；如果组成蛋白质的单体不同，象血红蛋白那样具有α链和β链者，则称异源多聚体。为了指明每个多体中的单体数目，通常使用的术语有：二聚体(包括同源或异源二聚体)、三聚体、四聚体、……八聚体等等 (图 163)。由二个或多个单体亚单位组成的蛋白质结构即称做蛋白质的四级结构(表 20)。

多肽链的一级结构规定着多肽链的二级结构，而一级结构、二级结构又进而规定着三级结构，所有这三种结构最终则决定了蛋白质的四级结构，对于这一基本认识人们已无异议。我们曾概括地说明在此种情况下，必定也会存在着某种合理的物理化学结构，作为将不同结构水平先后有序、由下而上联结起来的桥梁。但是

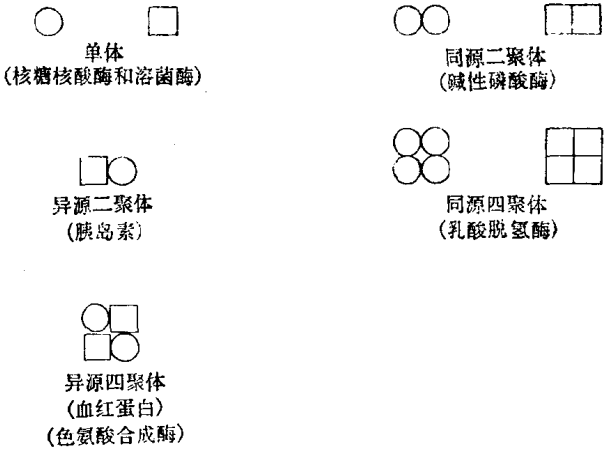


图 163　几种蛋白质的四级结构

正方形和环形代表单体亚单位，即单一的多肽链，每个多肽链均为折叠构象。单体存在着各种大小和形状，但多数接近球形。四级结构系产生于多聚体蛋白质中单体的聚合。(括弧中的大分子化合物分别代表各自相应的四级结构模型)

尽管如此，目前仍然无法直接从编码蛋白质一级结构的遗传信息去推断其四级结构或催化活性的本质。因此，有关四级结构的讨论亦只适用于探讨这样一些特定而具体的内容：1)遗传信息和蛋白质功能之间的关系，2)关于信息流在蛋白质各个结构等级水平上传递的动态(图 164)。

蛋白质的四级结构比蛋白质的三级结构更易使人理解。四级结构的本质仍不过是三级结构的一种功能。四级结构的最简单形式是一个多聚体蛋白质的形态，多聚体的形状一般依赖于 1) 单体的数目，2)单体的大小及形状，3)单体与单体间连接(结合)部位的特性。现已查明，蛋白质的四级结构同单体之间的聚合机制和这种聚合对单体本身构象产生的影响有关。另一方面，单体和多聚体的柔度（或称构象变动性）亦会影响它们的组合方式，而所有这些结构参量之间的相互关系则又有机地制约着蛋白质功能的表现。

大多数多聚体的四级结构可以在单体的二级结构和三级结构

表 20　一类由亚单位组成的蛋白质特征

蛋白质	分子量	亚单位	
		数目	分子量
胰岛素	11,466	2	5,733
β-乳球蛋白	35,000	2	17,500
血红蛋白	64,500	4	16,000
己糖激酶	96,000	4	24,000
乳房葡萄糖-6-磷酸脱氢酶	130,000	2	63,000
乳酸脱氢酶	150,000	4	35,000
延胡索酸酶	194,000	4	48,500
色氨酸酶	220,000	2	125,000
丙酮酸激酶	237,000	4	57,200
过氧化氢酶	250,000	4	60,000
乙酰乙酸脱羧酶	260,000	8	30,000
脲酶	483,000	6	83,000
β-半乳糖苷酶	520,000	4	130,000
	130,000	3～4	40,000
肌球蛋白	620,000	3	200,000
脊髓灰质炎病毒	5,500,000	130	27,000
丛缩病病毒	9,000,000	120	60,000
马铃薯病毒X	35,000,000	650	52,000
烟草花叶病毒	40,000,000	2,130	17,500

完全没有损害的情况下遭到破坏，这很可能是因为四级结构的丧失同单体的三级结构构象变动有关。此外，当多聚体解离成单体时，同样亦会使四级结构解体。多聚体的解离反应常常表现为可逆性，即单体能够重新聚合得以恢复多聚体原来的功能。在实验条件下，常常通过调节蛋白质溶液的 pH 值和离子强度来使多聚

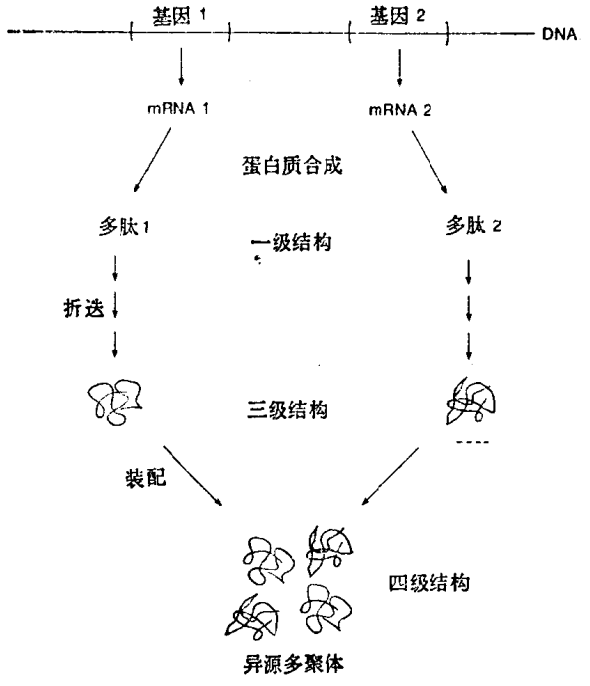


图 164　信息流在蛋白质各级结构水平上的传递

基因将信息传给多肽链一级结构。多肽链一级结构指导本身折叠成为具有三维结构的大分子。三级结构再进而使单体聚合，产生四级结构。酶在某些场合中起加速装配过程的作用，在另一些场合则表现为修饰某些氨基酸残基，被修饰的氨基酸残基反过来又可影响装配过程。

对折叠与去折叠反应所作的深入研究表明，三级结构只受多肽链一级结构的严格规定。例如，一个已被还原的(去折叠)处于变性状态下的(以尿素作为变性剂) 核糖核酸酶通过氧化作用可以部分复原成为某种三级结构的混合物。该混合物的酶活力只达天然核糖核酸酶活力的百分之一。但如果氧化条件温和，重新折叠回复的酶分子甚至也具有完全的酶活性(图 162)。此外，蛋白质在形成天然结构时所获得的构象自由能(free energy) 也会使折叠过程得到推动 (图 156)。现在已知还存在另一种大小为 42,000 道尔顿的酶分子，它能够加速折叠反应。这个酶并非专一性地只对核糖核酸酶的折叠反应行之有效，凡一切携带有接合不当的双硫键的多肽链，在该酶的作用下都概莫能外地会使之重新排成最为稳定的构象。不过一级结构和三级结构之间的对应关系并非总是铁板一块的。例如珠蛋白肽链、细胞色素c和脱氢酶在各个酶系内部均显示为相同的四级结构，其实这些酶蛋白多肽本身有着完全不同的一级结构。

体解离，在有些场合下可用某些变性药物，如尿素和盐酸胍(guanidine hydrochloride)等处理亦能奏效。有时移去变性药物后，虽然多肽链表面上仍然保持折叠如常的性能，但实际上单体的三级结构经过药物作用已不可能保持原样。因而在选用一种旨在使多聚体蛋白质解离的最佳方法时，工作中多是凭经验行事的。

在对蛋白质所作的离体实验中，多聚体的解离和重组是最为基本的现象。这一过程通常不可能在活体内观察得到。诚然，在活细胞里蛋白质很少象人工分离后那样处于溶液状态。已知许多酶都位于胞膜(线粒体膜，内质网和质膜)或核糖体等等胞器结构上，并且没有证据表明在活细胞中有任何一种酶是在溶液中行使功能的。因此在将离体的酶分子结构与体内的酶分子结构进行类比分析时必须谨慎从事。

假如某些特异型的单体在没有外来信息参与时能够自发聚合产生一类特异型的多聚体，则这个过程即被认为是一种自动装配。不需要外来信息意味着自动装配过程所必需的全部信息已经保留在单体本身的三级结构之中。比如说这个反应并不需要外来的酶所催化等等。现在看来，自动装配现象在生物界似甚普遍，可能大大超过需要经过多聚体酶催化的装配过程。前面已经讲过，核糖体的蛋白质亚单位就能被反复解离并重新装配成功能照旧的核糖体。此外，微生物纤毛、染色体组成、病毒颗粒等等也都具有类似的、靠预先形成的亚单位自动装配所产生的结构。

微生物的各种纤毛由一种或二种主要的多肽链单体组成。在功能性纤毛中，单体被装配成一个很大的(同酶的大小相比)多聚体管束状构造，这一装配过程一般均依赖于单体的三级结构（图165)。

另一方面，核糖体的结构方式更加复杂，其中含有50种以上的不同蛋白质还要再加上RNA大分子。尽管如此，多聚核糖体仍然可以在没有外来遗传信息参与的情况下挥洒自如地自动拆除和再装配成功(图166)。

核糖体的装配基本上是属于流水作业性质。可试举一例，30S

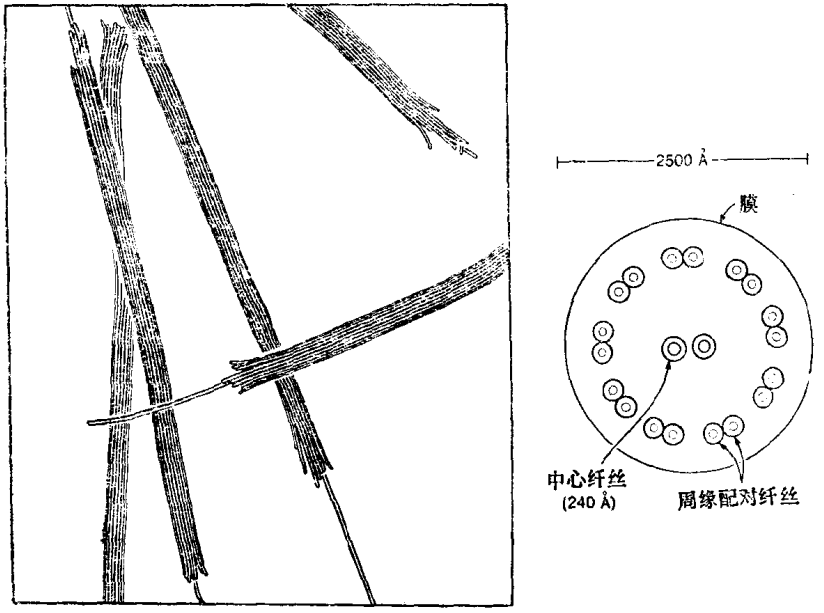


图 165 微生物鞭毛的自动装配

(a) 鞭毛虽然是一类比较复杂的细胞器，但基本上仅由一种称为鞭毛蛋白的单一的多肽亚单位所组成。装配过程从二聚体的形成开始，然后二聚体进一步聚合构成细长而复杂的纤丝。其复杂程度从图中体外装配过程中生长着的鞭毛端部即可见一斑。(引自 H. Koffler, 1965。)

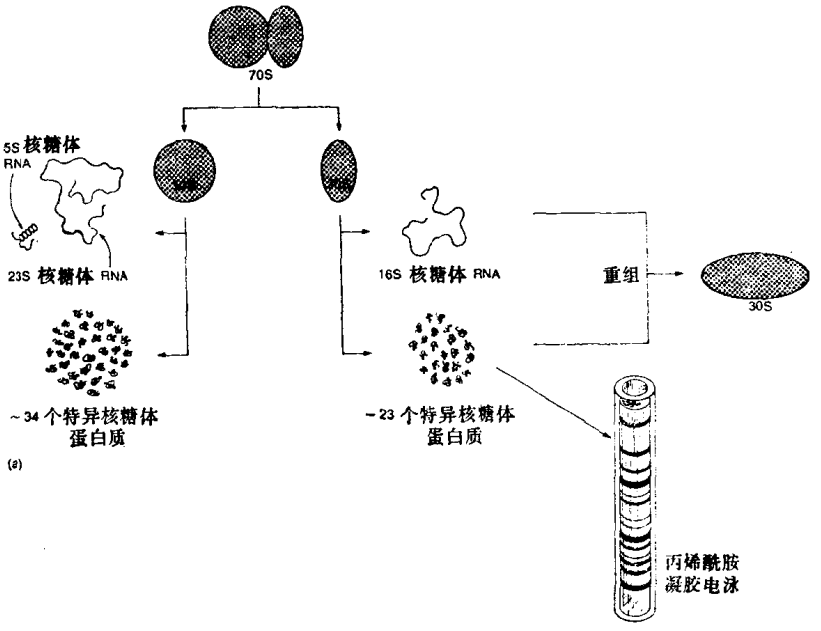
(b) 鞭毛管束结构微管横切面图。从横切面可以见到微管内部结构的复杂性以及作为纤丝组分的微管蛋白在聚合成长链时显示出的特有的排列方式。

核糖体亚单位由大约 23 个蛋白质和一个 16S RNA 分子组成，16S RNA 似乎位于核糖体内部而不为核糖核酸酶降解，这暗示 RNA 分子在核糖体蛋白质的装配过程中起一个结构单位的作用。最新实验表明，仅有少数核糖体蛋白质同 16S RNA 结合而其他蛋白质则不与之结合。然而，人们发现正是这些结合才启动了装配过程。而一旦装配开始，更多的蛋白质立即相继与之结合，使复合体成长发展，直到一个具备完整功能的 30S 亚单位被装配完成为止（图 166)。可以想象，如果每一种蛋白质的结合都能改变前面已装配了的复合体的空间构象以及改变各个组分蛋白质的构象性质从而

提供了新的结合部位，则上面这样一种装配过程便可以说是顺理成章的事情。

在详细探讨这一装配过程之前，应当记住一个由单体分子构成的酶识别其底物分子所必需的信息是由其本身的三级结构提供的。假如这些酶分子是位于胞膜或核糖体上，那么它们也必定含有决定其本身的定位和保持所据有的那个位置而必需的信息。换句话说，单体本身的形状就已经提供了许多特定的结合部位。在大多数场合，人们对于结合机制的本质不甚清楚，通常这类结合也都不要求共价键。一般认为，参与同其他单体进行聚合的单体分子均有其表面活性部位，这些部位规定了聚合的方式，随着也就决定了多聚体的整体形状。单体之间的结合一旦完成，每个单体的三级结构将会被修饰，从而产生出和(或)暴露出新的活性部位。现试举一例略加说明。

T_4 噬菌体是一个结构极为复杂的多聚体，其 DNA 大分子所载有的基因在 100 个以上并完全被包裹在一个由蛋白质亚单位组



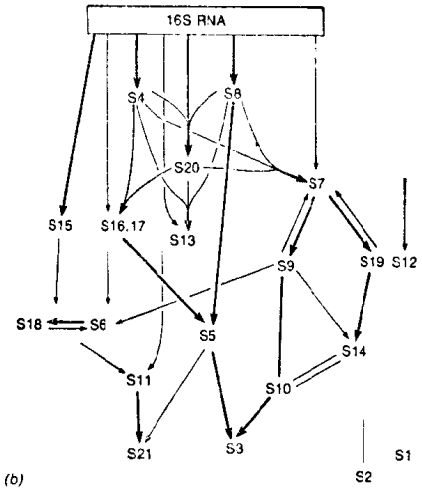


图 166　大肠杆菌核糖体亚基的组合和解离

(a) 大肠杆菌 70S核糖体可分为 50S和 30S两个亚单位。其中 30S亚单位又可分离成约 23 个不同的蛋白质(使用丙烯酰胺凝胶电泳法分离)和一个 16S RNA分子。将这些蛋白质同16S的RNA分子在37°C的KCl 溶液中混合之后，一个功能完全正常的30S核糖体亚单位即得以重建。从50S亚单位可解离得到一个 23S RNA分子和一个 5S RNA分子。(引自Watson, 1970。)

(b) 将从 30S 核糖体上分离出来的蛋白质记为 S蛋白质。这里 S 意指小的亚单位。符号后的号码表示各种不同蛋白质分子的编号。与拆卸过程相反，如将 30S 核糖体分离得来的各种蛋白质再回加到原先所结合在一起的 RNA分子上，每次只加入一种蛋白质，结果发现有些蛋白质分子直接同 RNA 分子结合。有些蛋白质则必须在前面的蛋白质已经与RNA分子成为复合体之后才能结合上去。还有少数种类蛋白质一定要等装配过程接近完成时方才结合。图中蛋白质之间的箭头表示一种蛋白质对另一种蛋白质-RNA复合体结合能力的依从性。例如，从S7指向 S19的箭头，即表示当S7-RAN复合体不存在时，S19不能直接与 RNA 分子结合。粗箭头与细箭头相比，前者的依从性更高。S12和S2这两种蛋白质要在复合体装配过程的晚期才能结合，其相依性不太明确。此外，对于 S11来说在结合前必须要求哪几个蛋白质与 RNA 结合还不清楚。(引自 M. Nomura, Sci, 179:869, 1973。)

成的多面体病毒头部里面。T_4 噬菌体除了上述头部结构以外，还具有一个外面围以伸缩性尾鞘的芯子所组成的尾部。尾部接在短颈上，短颈又同头部连接。尾部的末端装有六根突起的短尾钉和六根细长的尾丝，尾丝通过基板同尾部相接(图 43)。

此噬菌体的生命周史同样比较复杂，与图 59 中所示的T_2 噬

菌体不相上下。在感染时噬菌体通过尾丝和尾钉吸附在细菌的细胞壁上，然后尾鞘收缩推动尾部的芯子穿透细胞壁，头部载有的DNA分子随即经过芯子提供的通道进入细菌细胞内部。T_4噬菌体DNA进入寄主细胞之后，它所含有的信息立即引起细菌染色体解体，并控制寄主细胞中保留下来的代谢机能据以合成噬菌体自己的DNA分子和噬菌体蛋白质。最后，新合成的噬菌体DNA和噬菌体蛋白质再装配成新的噬菌体颗粒。一当几百个噬菌体颗粒装配完毕，细菌细胞即被噬溶，从而释放出大量新的病毒颗粒。

现已发现，噬菌体染色体在进入细菌细胞仅仅一分钟之后，细菌细胞中即便开始合成为噬菌体复制所必需的某些蛋白质。最早合成的蛋白质多为细菌病毒(即噬菌体——译注)DNA复制必不可少的各种酶，接着合成的是病毒头部蛋白质，最后合成的则是尾部蛋白质。图59所示即有关这些合成过程的一个总的时间顺序。当所有噬菌体的颗粒全部装配成功，便及时合成由噬菌体基因所编码其一级结构的溶菌酶，用以破坏寄主细胞的细胞壁，释放新的病毒颗粒。

正如任何能阻断必需代谢途径的突变会招致死亡一样，在噬菌体中如果发生一个突变使之不能合成某一部件，这样的突变也将表现为致死突变，称为生命周史不完全致死突变型(life-cycle lethals)。一般来说致死的代谢突变能够借助于外界补充该突变所缺失的代谢产物而得以保存，但这种方法并不能用来免除生命周史致死突变的致死效应。但是，假如借用某个突变型所产生的部件(尾丝，尾部，头部)，却有可能保存缺失这一类部件的突变型。为了行之有效地保存这一类生命周史不完全致死突变型，一般可以通过分离其中的条件致死突变型得到解决。如前所述，条件致死突变型在一定条件下表现致死，在另一种条件下不表现致死。在40℃条件下培养时，属于T_4生命周史不完全致死突变系统的某些突变型噬菌体只能发育到突变造成阻碍的那一步为止，嗣后即踏步不前，面临死灭。如把温度降到30℃，结果发现在某些突变型中，上述突变并不带来发育进程的阻断，噬菌体仍然顺利完成它们

的生命周史，从而使这一类突变型得以传种继代培养。

对于保存上述突变系统来说，还有另一种类似条件致死突变型的方法。这是采用二个寄主细菌品系，使生命周史不完全致死

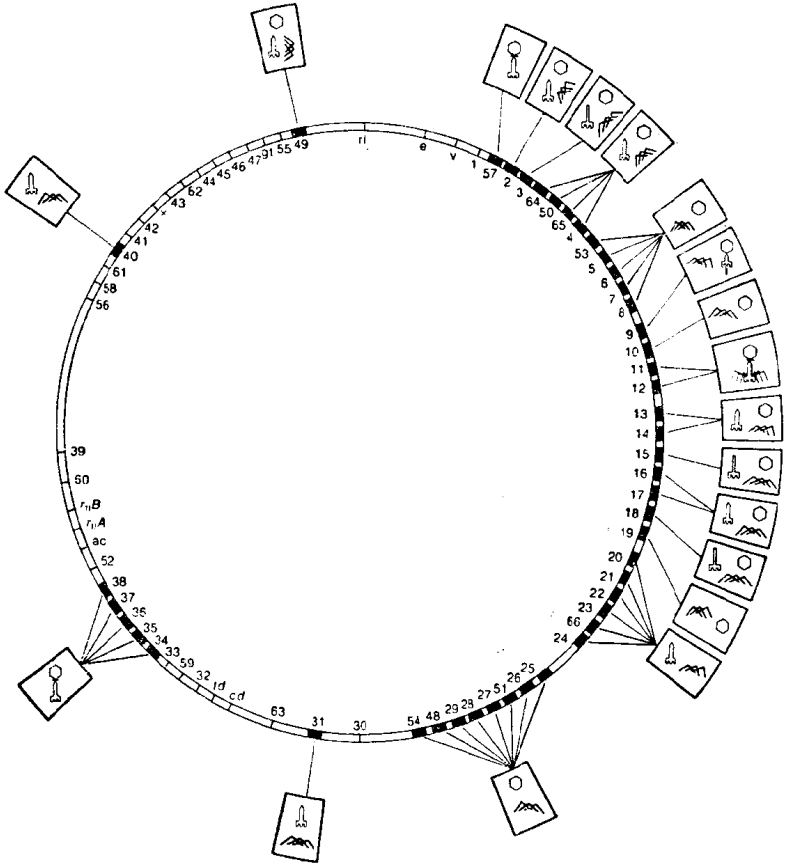


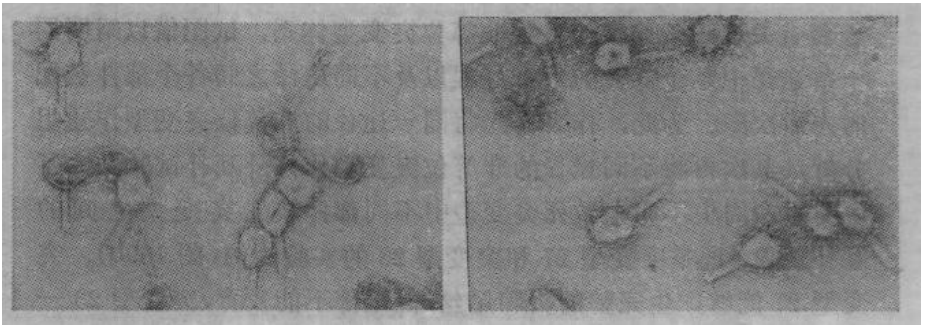
图 167　T_4 噬菌体的遗传学图

这张遗传学图只列出与形态发生有关的功能基因的位置。将生命周史不完全突变型噬菌体的各种类型感染寄主细胞，然后在电子显微镜下观察寄主细胞提取物，即可标出各类突变型所形成的形态结构部件。图上编号为各类突变型的突变位点并使之与各自对应的部件形成相联系。从图例可以见到在不同的提取物中，有些缺少头部，有些缺少尾部或缺少尾鞘或尾丝等等，花样繁多。在某些情况下，例如突变型 11 和 12，虽观察不到有明显的组分缺失，但这类突变型的病毒颗粒却无感染能力。上述所有各类有关形态发生的突变型可筛选其中的条件致死突变类型分离保存(如温度敏感)。图中基因号码旁边所标的部件缺失表型只有在限定的条件下才会表达。（引自 Wood 与 Edgar，1967。）

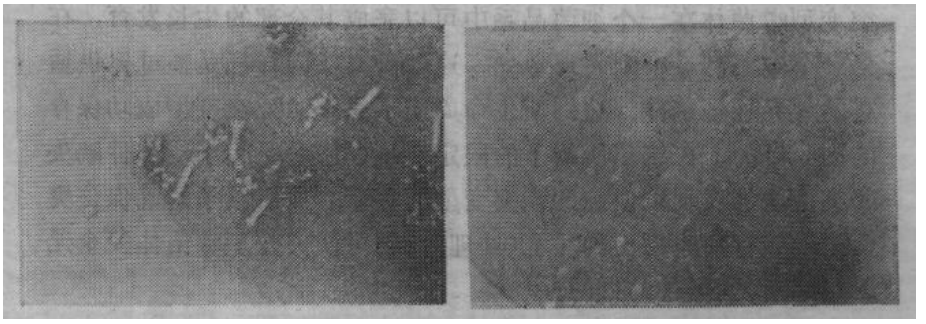
突变型噬菌体在一个细菌品系中可以完成其全部的生长发育，在另一个细菌品系中则表现致死。这二种方法同样都是通过提供适宜的外界环境条件，达到使生命周史不完全致死突变型成功保存的同一目的。在这里所给予的限定环境(处理)是为了使部件缺失的类型能够表现以便确证该突变型的存在。而所给予的确保该突变型能够存活的另一种环境(处理)，则可使突变型噬菌体安全无恙，从而得到保种的目的。

生长在适合的寄主或者适宜的温度下的生命周史不完全突变系统中的各种突变型，可以使用常规的遗传分析方法进行研究：例如为了研究噬菌体染色体上突变型的连锁关系，可将不同类型的生命周史不完全致死突变型作杂交试验（图 167)。由于一个或多个部件的缺失或者一个或多个为装配程序所必需的酶缺失，均能造成噬菌体发育阻断，所以当把这些同一缺失表型的不同突变置于一个限定的环境(或某一寄主品系)中培养时，即可诊断不同突变型噬菌体所可能达到的发育阶段。

借助于电子显微镜，目前已有可能识别为数甚多的生命周史不完全致死突变型以及感染后在致死培养环境下突变型中有关各种部件的缺失图景（图 168)。将图 168 同图 167 比较后，便能明白无误地说明基因 23、31 和 40 为形成病毒头部所必需，而基因 19、10、6 和 27 则为病毒尾部形成所必需。从 T_4 整体装配的总的图景来看噬菌体的个体发育，大体具有三条主要分支：第一条途径



(a)　　(b)



(c) (d)

图 168 未装配的噬菌体组分的电子显微镜照片

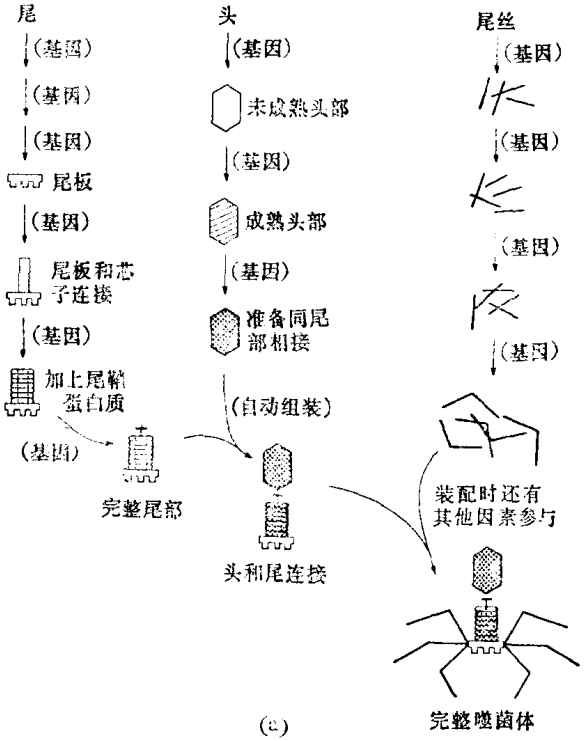
(a) 装配完好的正常 T_4 噬菌体。
(b) 缺少尾丝(未装上)的噬菌体。在图片背景中可见到少数未组装的尾丝。
(c) 未装配的尾部。摄自由头部缺失突变型所感染的细菌溶菌液。
(d) 采用蔗糖梯度离心法收集到的游离基板。摄自尾管形成缺失型噬菌体感染的细菌溶菌液。(照片由Jonathan King 与 Elaine V. Lenk, MIT 赠送。)

产生头部,第二条途径产生尾部,第三条途径产生尾丝。在这里一条途径中的某种突变阻碍并不妨碍其他二条途径所控制的组成部件的形成,但是最后的装配成功仍必需依赖于所有三套组成部件的配套齐全(图 169a)。

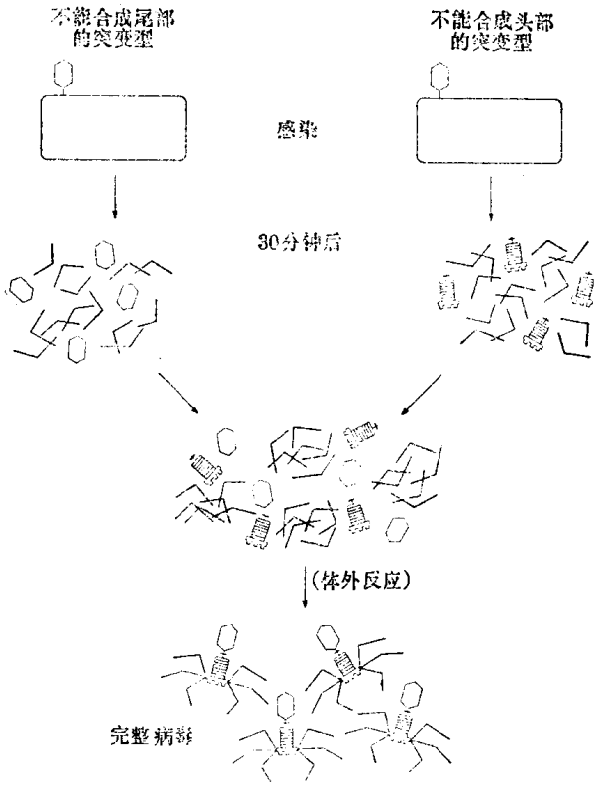
在决定头部、尾部和尾丝形成的三条途径中,每一条途径内部还存在有种种发育障碍,这可以从每一条代谢途径中都表现有许多非等位突变型的事实来判断。Robert Edgar 曾成功地建立了一套有关这三条途径的各种非等位突变型体系,试图借以阐明同一条途径中某个部件的发育过程以及不同途径之间各个部件装配的先后次序。为此,作者首先将同一途径的非等位突变型作重组分析,其次再将不同途径的非等位突变型作体外互补试验。为了说明生命周史不完全致死突变型中不同部件缺失突变型之间的互补现象,现举突变型 27 和突变型 23 的实验为例(图 169b)。突变型 27 能够产生完整的头和尾丝,但尾部不能发育,突变型 23 一般在限定的环境中产生尾部和尾丝,但缺失头部。当把由前者分离得来的头部与后者分离得来的尾部及尾丝混合培养(存在有突

变型 23 感染过的细胞的酶提取液)时，便会立刻看到感染力完全的噬菌体颗粒迅速增生。电子显微镜图片表明，所有这些噬菌体颗粒大体都具有正常的完整无缺的外部形态。

根据采用体外互补或体外装配的方法，对所有的发育突变型(即部件缺失突变型)进行分析，作者提出了一个能够据以准确定出有关自动装配进程中若干发育步骤的模型。这些步骤可概括地说明三条代谢途径各自的发展顺序，同时也能够辨别出自动装配过程中哪些突变型在部件发育的哪一步骤上能通过。在这里，有一些早期的装配步骤和最后一个装配步骤大概需要有酶进行催化，在这些步骤上阻断的突变型相互之间并不是部件对号就能够互补的，即不表现为简单的互补。根据不同的突变型能够形成的部件和所缺失的部件(或形成功能不全的部件)之间的差别，可以



(c)



(b)

图 169 感染力正常的 T_4 噬菌体的形态发生程序

(a) 在形态发生过程中，每一步骤均由基因指导。箭头代表装配过程进行的顺序。(引自 Wood 与 Edgar, 1967。)

(b) 互补现象可以在某些情况中得到证实。例如，若将突变型 27 和 23 的提取物在体外混合，则具感染力的完整噬菌体颗粒便可以再生。(引自 Wood 与 Edgar, 1967。)

划归不同的互补群，目前鉴定的互补群至少已有 13 个之多（表 21）。

在一个个按正确顺序鉴定三条代谢途径的各自发育步骤时，人们遇到的另一个困难是电子显微镜的分辨力不够敏锐。往往对于极其相近的突变型，在判别其表型之间的差别方面无能为力。实际上，性状发育的每一个步骤可以小到只包括形态上的微细变化，

表 21　体外装配实验中 T_4 噬菌体突变型互补群

提取物编组	突变基因	能够形成的部件	缺失部件或功能不全部件
I	5,6,7,8,10 25,26	头,尾丝	尾
II	20,21,22,23 24,31	尾,尾丝	头 (偶尔可形成无功能的头部)
	2, 4, 16,17 24,50,64,65	头,尾,尾丝	头部虽有若无 (全部部件虽有若无)
III	54	头,尾丝	尾芯
IV	13,14	头,尾,尾丝	?
V	15	头,尾芯,尾丝	尾鞘
VI	18	头,尾芯,尾丝	尾鞘
VII	9	头,缩小的尾部,尾丝	?
VIII	11	无感染力的完整颗粒	?
IX	12	无感染力的完整颗粒	?
X	37,38	头和尾连接	尾丝
XI	36	头和尾连接	尾丝
XII	35	头和尾连接	尾丝
XIII	34	头和尾连接	尾丝

而这却是电子显微镜所不能觉察到的。例如在电镜图片上，虽然见到似乎已经充分发育的完好头部，但就是不能同完好的尾部组合装配，即为明证。

在T_4病毒颗粒的装配全过程中，最后的一步是把尾丝接到头-尾复合物上。可是当缺少某一种酶时，这一步即无法通过。这种酶存在于尾丝提取物中，在不能形成尾丝的突变型提取物中以及在未被 T_4 噬菌体感染过的大肠杆菌提取物中都从未发现过这种酶。因此，看来这种酶大概是由细菌病毒基因编码，在合成上也是同尾丝的发育相联系的。

T_4 噬菌体的发育有其顺序性，个体发育的大多数步骤已包含在预先合成的亚单位的自我装配过程之中。我们虽不能根据噬菌

体的发育体制来推论那些更为复杂的生物种属的发育系统，但它却是一个证据，说明预成式的亚单位的自动装配乃是一种普遍的生物学现象。目前，对于有关的遗传信息究竟怎样通过蛋白质的一级结构指导各个亚基组装为完整的噬菌体颗粒、核糖体，以及酶系统如何与胞膜的特定部位相结合等等许多引人入胜的问题虽然还作不出明确的回答，但是正象对发育含义的一般理解不断提高一样，人们对自动装配重要性的认识势必也会进一步提高。事实上，自动装配过程已经反映出遗传信息传递中的一个步骤，要想确切弄清信息传递的详细机制恐怕还需要一段时间。

人们相信，多聚体酶蛋白的自动装配过程包含着大体类似于某些较复杂的多聚体装配系统中所存在的化学机制。由于这一认识，再加上酶的结构又比较简单(同 T_4 噬菌体相比)，因此酶分子往往成为比其他任何信息大分子更为热门的研究材料。过去酶学研究收集了大量有用的数据并建立了不少模型，不过在结构与功能的关系和有关酶的一般理论方面仍然混乱不堪。下面讨论的酶催化作用机制仅仅是此类浩瀚文献中的海滩一贝，所用材料也是根据与信息流主题的关系选定的。有关酶学和催化作用方面更为详尽的资料，可查阅文末参考文献。

酶的作用机理

蛋白质的组成文字共含有 20 个常用词汇，即 20 种氨基酸。大多数的蛋白质由 50 个到 750 个氨基酸残基组成，但即便是含 50 个氨基酸，其可能的排列组合数目(亦即构成蛋白质的种类数)也还是个天文数字(达到 20^{50})，这要比维持当今全部生物种属在生命活动中行使代谢和生理诸方面功能所必需的蛋白质种类的总数还多得多。从目前所存留的植物、动物和原生生物来看，其生命所需蛋白质的氨基酸排列顺序似乎有某种选择性；或者说迄今为止在所有蛋白质中发现的氨基酸不同排列顺序的总数仅仅局限于可能组合数的一个极小范围，从而表现出高度的统一性。可是从另一方面看，要说在过去的 30 亿年间自然界曾经对氨基酸的各种可

能组合无一遗漏地进行过选择，当然亦难以令人置信。因此，除了要探求一级结构的顺序是怎样对三级结构和四级结构的性质和功能发挥影响之外，还必须阐明为什么在可能的氨基酸组合总数中对生物系统至关重要的只有这样如此之少的一个份额？这些特殊的范围极其有限的组合又是怎样和为什么能够在进化过程中被卓有成效地选择留存下来？这方面的内容拟在下一节细胞色素 *c* 蛋白质——广泛分布于系统进化树中不同生物种体内的一种蛋白质——的讨论中详加论述。

从分子遗传学的现代概念来看，大体可以断定折叠蛋白质的表面具有一个以上（数量和种类）的特殊结合（活性）部位。另外，许多蛋白质除了有识别底物分子的部位之外，还能够识别调节物质和其他能与之结合的各种大小不一的分子，这些分子同酶结合之后可在很大的范围内改变催化速率。因之，至少在有些酶的表面区域对于调节反应速率以及指导终止的简单信号和启动催化反应的多种信号是表现敏感的。下面我们就要来讨论这个问题。

毋庸置疑，表面活性部位决不会是僵死的不可变动的铁板结构，事实上正是活性部位的灵活性才反映出蛋白质三级结构和四级结构的可变性。当然，也有理由这样设想：即三级和四级结构的变动可能反而是促使活性部位变动的推动力之一。在某些酶中，底物的结合部位小到甚至只有三个氨基酸残基。这些活性部位上的残基在一级结构中常常彼此隔开很远，通过三级结构的作用才将它们定位在邻近的具有活性的位置上，四级结构则为活性部位的灵活变动提供了推动力。

蛋白质（至少是催化蛋白质）的四级结构可能是在三级结构固有的功能基础上作为某种"改进"并经历了漫长的进化历程，逐步衍变产生出来的。换句话说，即四级结构能以更为有效的和多方面适应的方式来识别调控信号并作出反应，同时还为催化作用的活性部位提供了构型变动的可能性。一般认为，指导四级结构形成的信息是自我保存在参与组成单位的三级结构之中，而三级结构的信息又可以从其一级结构追本溯源。迄今为止，对于这一说

法尚无足够的理由可据以提出怀疑。所以我们不妨带着这一先成之见来详细分析一下酶分子的催化功能，这类催化功能现已能采用多种不同的技术成功地给予测定(表 22)。

表 22　取决于酶促反应性质的各种酶活力的测定法*

A. 直接法

1. 流体压力计测量法(测定成分为气体的场合)

a. 耗氧量(氧化酶活力测定)

b. 二氧化碳排出量(脱羧酶活力测定)

2. 层析法

a. 通过纸层析计算底物含量，例如采用放射性等

3. 化学测定

a. 计算消耗的无机磷酸(磷酸酶，磷酸化酶活力测定)

B. 间接法

1. 分光光度法

a. 脱氢酶的活力测定(当一种底物的吸收光谱不同于另一种底物时)

b. 脱色反应，例如使甲基蓝或 6-二氯靛酚变色

2. 电极测定法(用于酸度增加的反应)

a. 记录 pH 值的变化或记录为保持恒定的 pH 需加入的试液量

* 活力单位(适用于任何一种酶)等于每分钟催化 1 微克分子底物转化所需的酶量(应有过剩底物存在)。还有另一种特别的活力单位规定，即每毫克蛋白质所具有的活力单位数(比活性)。

大多数酶含有 50 个到 750 个氨基酸残基，而组成多聚体酶的每个单体各有一个同底物结合的部位。参与组成结合部位的氨基酸残基数目同全部残基总数相比是微不足道的。确定酶的表面活性部位以及鉴定活性部位的氨基酸残基是酶学实验一直关注的问题之一。通常用来鉴别活性部位残基的方法有两种。

第一种方法是，使用不同种类的具有特殊反应基团的化学物质与活性部位的氨基酸残基复合，以掩蔽活性部位。例如，将活性部位上的赖氨酸羟化即可导致酶的失活；此外，如果活性部位上的

丝氨酸与二异丙酯氟磷酸(DIPFP)复合，或者半胱氨酸与诸如 P-氯苯甲酸汞盐等汞制剂复合，一般都同样可使酶失活。

第二种方法是以突变为手段来替换活性部位原有的氨基酸残基，以期建立一个表现不同变化的酶活性突变系列。从理论上说，对于所产生的这些活性变化可以从每个外来替换的氨基酸残基所给予活性部位的影响来检测和说明。

上述两种方法都会因为下列原因受到限制，即氨基酸置换可以发生在整个蛋白质分子上的任何部位，而且亦未必不招致催化性质的改变。这一事实一方面揭示了蛋白质整体结构与其活性部位结构之间的统一性，但另一方面，这也使得旨在精确决定活性部位上实际上是哪一些氨基酸残基的问题变得更为困难。读者只要回顾图 134，就会见到在整个色氨酸合成酶的 α 多肽链上，不同位置上的氨基酸残基的替换都能改变色氨酸合成酶的催化活性。

幸而现已找到了一些对活性部位具有专一性的，仅只同位于活性部位上的氨基酸残基复合的药剂，二异丙酯氟磷酸(DIPFP)即为其中一种。DIPFP 是许多酶的有效抑制剂，它不仅专一性地与丝氨酸残基起反应，而且还有选择性地同一定种类的酶活性部位上丝氨酸残基起反应。因此，在检测某个酶的活性部位时，可使用 DIPFP 先与活性部位上的丝氨酸残基复合，然后，将此酶水解以鉴定是那段肽片上的丝氨酸同 DIPFP 产生复合。如果该酶的一级结构预先已知，则活性部位上丝氨酸残基所处那段肽段在酶的结构上的位置便可据以确定，并进而可能去说明那些位于活性部位上的氨基酸残基所具有的特异的或独特的性质。在这方面，它们同多肽链上其他氨基酸残基应有不同。

如果酶的一级结构已经知道，该酶活性部位上的丝氨酸残基又予鉴定，那么即可查明同活性部位上丝氨酸残基相连接的其他残基。非常有趣的是，人们发现在大多数酯酶和蛋白酶中与活性部位上丝氨酸残基连接的残基彼此雷同(表 23)。除此之外，通过改进技术手段在识别活性部位氨基酸残基方面亦可达到异曲同工之妙。办法是先使酶同底物结合达至饱和，借此阻止同特定氨基

表 23　活性部位含有丝氨酸的肽段

酶	来　源	肽　顺　序
羧酸酯酶	马肝脏	-甘-谷-丝-丙-甘-甘-
乙酰胆碱酯酶	电鳐电器管	-谷-丝-丙-
拟胆碱酯酶	马血浆	-苯-甘-谷-丝-丙-甘(丙$_2$, 丝)-
碱性磷酸酶	粘质沙雷氏菌	-(苏或丝)天冬-丝-丙-
碱性磷酸酶	牛　肠	-天冬-丝-丙-
胰蛋白酶原	牛胰腺	-半胱-谷酰-甘-甘-天冬-丝-甘-甘-脯-缬-缬-半胱-
	猪胰腺	-甘-谷酰-甘-甘-天冬-丝-甘-脯-缬-缬-半胱-甘-
胰凝乳蛋白酶原 A	牛胰腺	-丝-半胱-甲-甘-天冬-丝-甘-甘-脯-亮-缬-半胱-
胰凝乳蛋白酶原 B	牛胰腺	-丝-半胱-甲-甘-天冬-丝-甘-甘-脯-亮-缬-半胱-
弹性蛋白酶	猪胰腺	-半胱-谷酰-甘-天冬-丝-甘-甘-脯-亮-组-半胱-
凝血酶	牛血浆	-甘-天冬-丝-甘(谷,丙)-
枯草杆菌蛋白酶, Nagarse	*B. subtilis*	-丙-苏-天酰-甘-苏-丝-甲-丙-丝-脯-组-
枯草杆菌蛋白酶, Carlsberg	*B. subtilis*	-苏-亮-天酰-甘-苏-丝-甲-丙-丝-脯-组-
曲霉肽酶	*Aspergillus oryzae*	-苏-丝-甲-丙-

表中所列为几种含有丝氨酸的酯酶和蛋白酶及其活性部位中与丝氨酸残基邻接的氨基酸残基顺序。

酸起反应的药剂与活性部位的残基复合。这样当酶的肽链上其余氨基酸全部同某一特效药剂复合时，那部分因与酶饱和而未起反应的活性部位残基便得以保存下来供鉴定之用。在定位过程中可以把各种方法相互结合起来，按这一方式所查明的由有关氨基酸组成的活性部位目前已不在少数(表 23)。

从这些研究可概括如下二点：第一，虽然活性部位上氨基酸残

基的数目为数甚少，但是催化过程乃是由许多其他残基同活性部位残基的协同作用。协作氨基酸残基在多肽链一级结构中彼此相隔颇远，意味着多肽链的折叠构象对催化作用是至关重要的。其次，这些研究还为理解这一问题提供了线索：为什么整个酶分子同其活性部位相比一定要大许多。为什么在分析每个活性部位时单单由其一级结构顺序来推导活性部位的功能是远远不够的（图134和表23）。

酶学研究指出，酶的催化作用常常依赖于称为辅助因子（辅酶或辅基）的存在和参与。辅助因子包括好几类，金属离子是一类，例如锌、钴、钙、铜、铁、钾、镁、锰和钠等等；维生素是另一类，包括维生素B_{12}、核黄素和烟酸；第三类辅助因子为有机大分子，只存在于某些酶中，主要是脂类和碳水化合物。在大多数情况下，酶促反应中辅助因子和酶蛋白之间的相互协作机制虽不清楚，但知道它们对酶的活性却是不可缺少的。有些场合，辅助因子同酶蛋白牢固地结合，但是两者以共价键形式结合者颇为少见。通常采用控制pH值和离子强度的方法即可轻而易举地把大多数辅助因子从酶中析离出来。

金属离子是最常见的辅助因子，但是人们对其在催化过程中的作用机制很少了解。在某些场合（例 Cu-抗坏血酸氧化酶复合物），金属离子辅助因子直接与酶结合。在另外的场合，往往是金属离子先同底物结合，然后再被酶识别。

有一段时间里，人们曾相信生物体系内维生素的唯一功能是作为辅酶在起作用。营养成分中对维生素的需求量非常之少，即只要有微量的维生素分子即能满足生命需要这一事实，与上述观点亦不谋而合。不过，现在已有证据表明，有一类维生素可能也具有其他方面的某些功能。

许多酶只有在有机大分子辅因子存在的条件下才赋有功能（表24）。迄今人们仅仅知道辅因子是催化反应不可缺少的因素，例如琥珀酸脱氢酶一旦脱去同其结合的脂类分子，此酶便毫无活力。加入所脱去的脂类分子，酶活力又能得以恢复。关于催化作

表 24　不同酶的辅助因子(或辅酶)及其功能

辅助因子	酶功能
有机分子辅基类	
ATP	转磷酸(作用)
抗坏血酸	羟　化
生物素	CO_2转移
辅酶A(CoA)	脂肪酸合成和氧化
核黄素	携带氢原子
铁原卟啉	过氧化氢酶,细胞色素蛋白和珠蛋白辅基
烟酰胺腺嘌呤二核苷酸	携带氢原子
磷酸吡哆醛	氨基转移酶和脱羧酶辅酶
四氢叶酸	转移一碳物
硫胺素焦磷酸	氧化脱羧
离子激活剂	
钙	ATP 酶
钴	需要维生素 B_{12} 的酶
铜	细胞色素氧化酶
锰	多聚酶
镁	肝脏精氨酸酶
钼	金属黄素蛋白
锌	碱性磷酸酶
铁	氧化酶类

用和辅因子之间的关系问题,这里暂且搁置不议,因为这并不妨碍下面要讨论的溶菌酶这一主题。

溶菌酶是一类比较小的蛋白质,它催化着如图 170 所示的反应。图 171 是根据 X-衍射图所确定的有关这一蛋白质的三级结

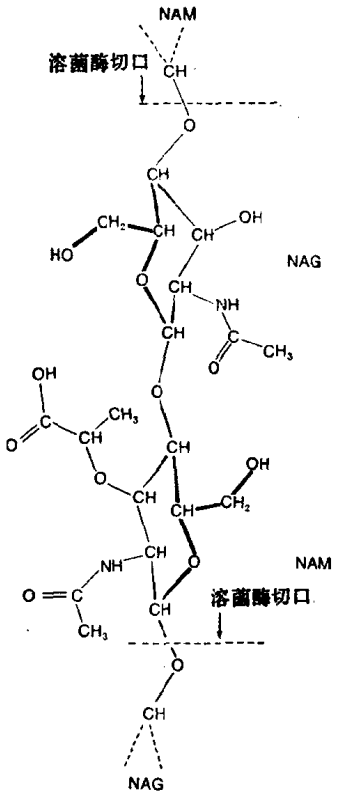


图 170　溶菌酶的催化反应

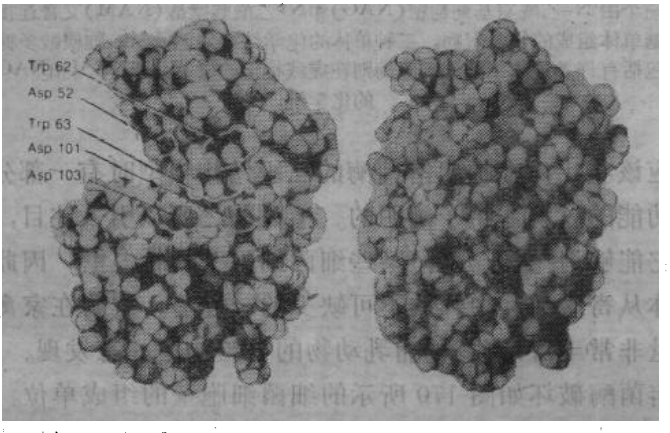
溶菌酶通过分解细菌胞壁中胞壁质的多糖结构组分而起溶菌作用。胞壁质含有一个由 N-乙酰氨基葡萄糖(NAG)和N-乙酰胞壁酸(NAM)交替连接的氨基糖单体组成的长多聚物。二种单体的化学结构如图所示。胞质酸多糖的主链包括有许多单体单位。溶菌酶则在虚线标记的位置切断NAM和NAG之间的化学键。

构。应该指出，在提出溶菌酶的三级结构模型时有一部分是通过对其功能的了解才给予阐明的。溶菌酶之所以引人注目，最早是因为它能够消化(或分解)某些细菌细胞壁的粘多糖。因此，它是噬菌体从寄主细胞中释放不可缺少的一种酶。此酶在家禽的卵白中含量非常丰富，在多种哺乳动物的分泌物中亦有发现。

溶菌酶破坏如图 170 所示的细菌细胞壁的组成单位。这种单位包含着本身以 C-1→C-4 方式连接的由己糖组成的多糖，因而

具有纤维素和细菌细胞壁多糖的特征。多聚物由 N-乙酰氨基葡萄糖(NAG)和N-乙酰胞壁酸(NAM)二种单体交替组成。溶菌酶则在NAM的C-1 和与之相连的位于NAG 上的氧原子之间将糖苷键切断。

在体外系统中,溶菌酶可识别作为底物的多聚-NAG 分子,却不能识别多聚-NAM 分子。当此酶与 NAG 三联体分子(即由三分子的 NAG 组成——译注)混合培养时,活性部位即同三联体分子形成稳定的复合体,从而丧失进一步识别底物的能力。也即是说,该酶所催化的反应不可能再行发生。曾使用X-衍射方法对溶菌酶-NAG 复合物的结晶体及裸露的溶菌酶(即未与 NAG 复合的溶菌酶——译注)的结晶体空间构型作了比较。结果发现,当溶菌酶未与 NAG 复合时,该酶的三级结构接近球形,大体横贯其最短的直径切开一个大的凹槽(豁口),在槽的一侧是螺旋结构(残基顺序在 1 至 40 和 101 至 129 之间)占优势的区域,槽的另一侧构型不太规则(残基顺序从 41 至 87),把豁槽两侧连在一起的是另外一个螺旋片段(残基顺序从 88 至 100)。从整体上看,因为只具有三个规则的螺旋区段(5 至 15,24 至 34 及 88 至 100),所以溶菌酶的 α-螺旋构型比之肌红蛋白为数更少,加上从残基 24 到 34 和从 88 到 100 这两个区段呈旋转状,结果使构型显得更不规则(比



(a)

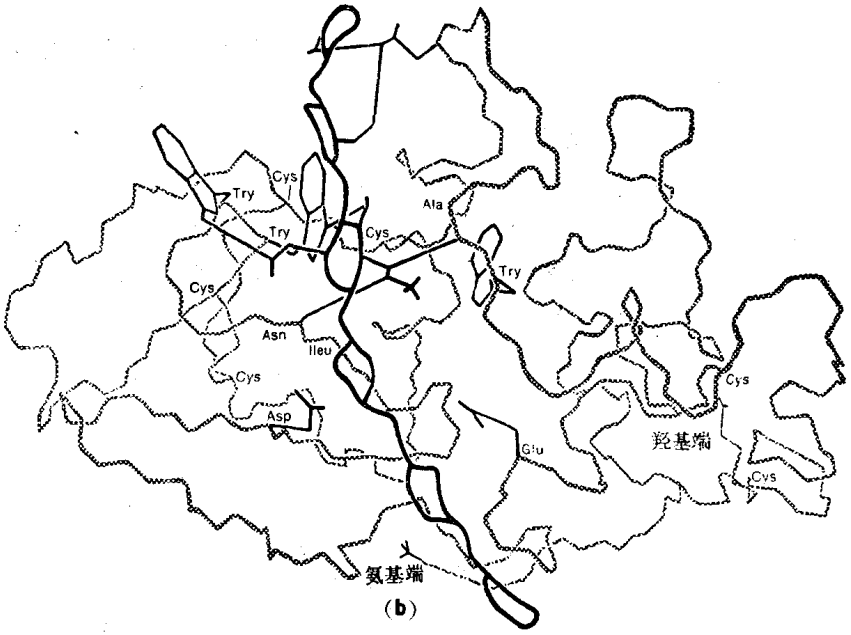


图 171 溶菌酶的三级结构(溶菌酶三级结构的二种模型)

(a) 溶菌酶与底物结合时，豁槽嵌满；在不与底物结合时，豁槽敞开。这二个模型真实地展现了原子间的空间关系，这就是一般所谓嵌合补缺模型。活性部位构成的豁槽极其精巧，其"接纳"底物分子的本领达到了天衣无缝的程度。沿豁槽排列的是几个特殊的能起催化作用的氨基酸侧链，与底物结合后催化水解反应。

(b) 本图以许多折线和闭环表示溶菌酶多肽主链的走向同被结合的底物分子之间位置关系。粗线代表被结合的底物分子和溶菌酶中起催化作用的氨基酸残基。(引自 Dickerson 与 Geis, 1969, 及 D. C. Phillips, 1966。)

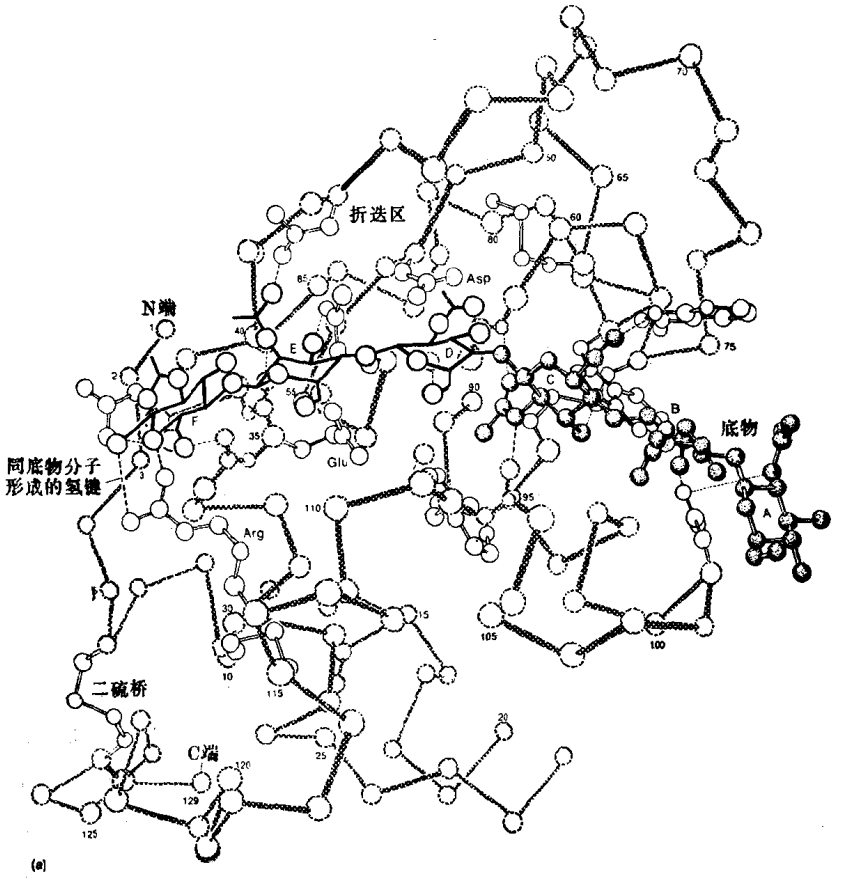
较图 160 和图 172)。

无论是与底物复合的溶菌酶，或是未与底物复合的溶菌酶，都具有同样的三级结构。前面的 40 个氨基酸残基以二个长度不同的α-螺旋和一个β 构象形成一个致密的单位。往下八个残基(从 41 至 48)呈直线排列，随后链反向回折五个残基(从 49 至 54) 并与残基 41 至 48 的区段反向平行从而构成并列双链。残基 55 和 56 是二个非极性氨基酸，正好位于豁槽中段的底部；从残基 57 开始，多肽链又反转回折直至残基 60。由于中间发生了二次反转，所以从残基 41 至 60 形成了一个"之"字形的三链式并列结构。残基

61 与并列三链垂直向外伸展(图 172)。

溶菌酶的三级结构与肌红蛋白亦有相同之处，即分子内部多数为非极性氨基酸残基，外部多为亲水性残基。除二个特殊的残基之外，几乎所有的亲水基团都分布于分子表面。多肽前面 40 个残基中的二个螺旋区段均含有非极性残基，在三级结构中彼此紧密靠接。从残基 88 到 100 的螺旋区段有一个非极性残基埋在豁槽的底部，正好将溶菌酶分成二半，形成一个可以嵌入底物分子的开口。而 β 折叠则整个地由非极性残基组成。

三体-NAG "底物"同溶菌酶的豁槽相结合，结合面约为豁槽



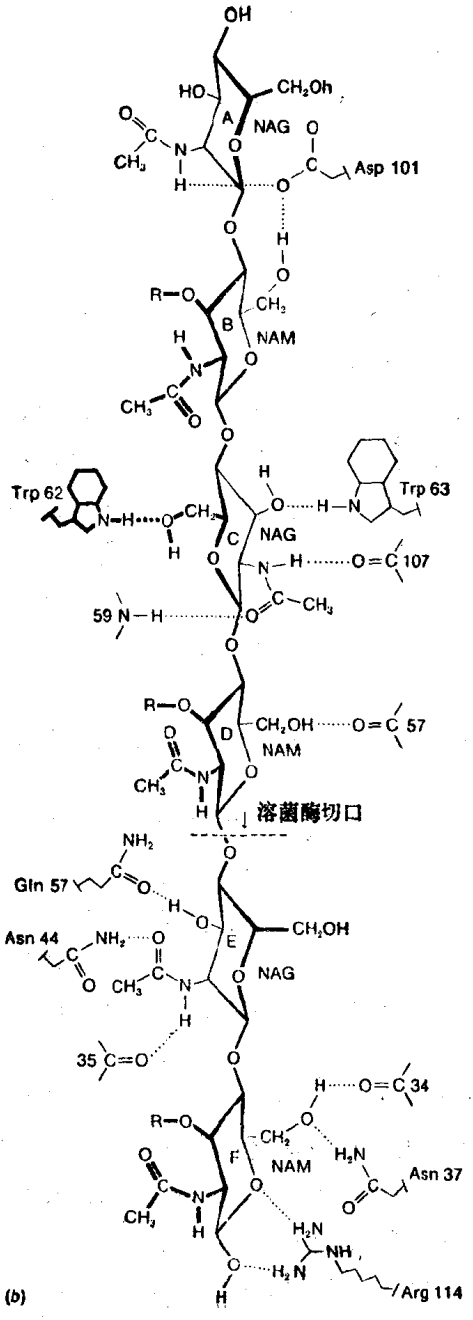


图 172 活性部位与底物结合的溶菌酶（来源于家鸡）三级结构

(a) 图示底物分子水平地横贯于溶菌酶分子的折叠构象之中。α-碳原子的位置用数字标明，字母A～F则表示底物分子的环状部份。(引自 Dickerson 与 Geis, 1969。)

(b) 多肽链氨基酸残基和形成活性部位有关。三级结构导致了活性部位的氨基酸同底物接触，即一级结构中彼此相隔较远的氨基酸残基在折叠构象的活性部位中相互接近。氨基酸的编号与它们在一级结构中N端起始的位置相对应。(引自 Dickerson与 Geis, 1969。)

长度的一半。这暗示，此处应是作为活性部位的场所。当底物分子存在时，在此区域能观察到三级结构会出现微小变化：例如第62个Trp残基此时向底物的B环移动约0.75Å，而且整个槽面亦随之向底物靠近。从现有的结构参数估算，溶菌酶分子的豁槽宽度大概正巧可以塞进一个己糖单元。图173所列即为据以设想的有关这一催化过程的程序。

另外，核糖核酸酶和溶菌酶也至少有一个共同的特征，二者均可识别和切断长链的底物分子。X-衍射证明，核糖核酸酶分子的N-端含有二个短小的以其非极性残基相互邻接并朝向分子内部的α螺旋区域。核糖核酸酶也有一条引人注目的沿酶分子长轴方向曲折延伸的裂缝，双链β构象贯穿其中。使核糖核酸酶同抑制剂2′,3′-环状尿嘧啶结合形成稳定的复合物，再水解分离鉴定。结果表明，核糖核酸酶的活性部位氨基酸残基为：His 12, His 119和Lys 41。这与X-衍射得出的结论相一致(图174)，即His12, His 119和Lys 41等氨基酸位于同抑制物结合的裂缝里，彼此紧密相邻。

关于蛋白质的空间构象及其活性部位氨基酸残基提供配位基的问题，从其他结晶的X-衍射研究也都充分说明在酶的构象结构和催化功能之间有着一个不可分割的联系。不过这些研究并未指明。那些指导一级结构转变成为三级结构的作用力性质，也未说明一旦三级结构形成之后稳定这种折叠构象的作用力本源。我们只能从溶菌酶和核糖核酸酶的结构模型看到，蛋白质的三级结构对于提供和维持酶的活性部位的功能是一个不可缺少的基本因素。

在不含抑制物和供给充足底物的试管条件下，酶的催化反应速率在底物耗尽或者产物浓度引起负反应之前是始终如一的。在控制的条件下每次加入一种诸如抑制物、辅因子、人造底物、复合药剂等等物质，并根据活力速率变化测量酶促效应。结果发现，测量速率变化无论怎样精确，试验设计方法无论怎样多种，要想将试管环境模拟为活体内部的酶催化过程总是不尽如理想。在复杂的

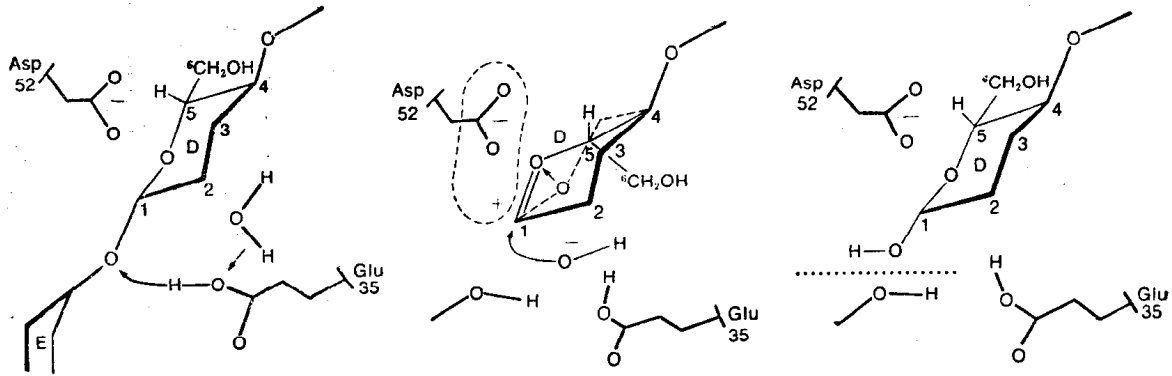
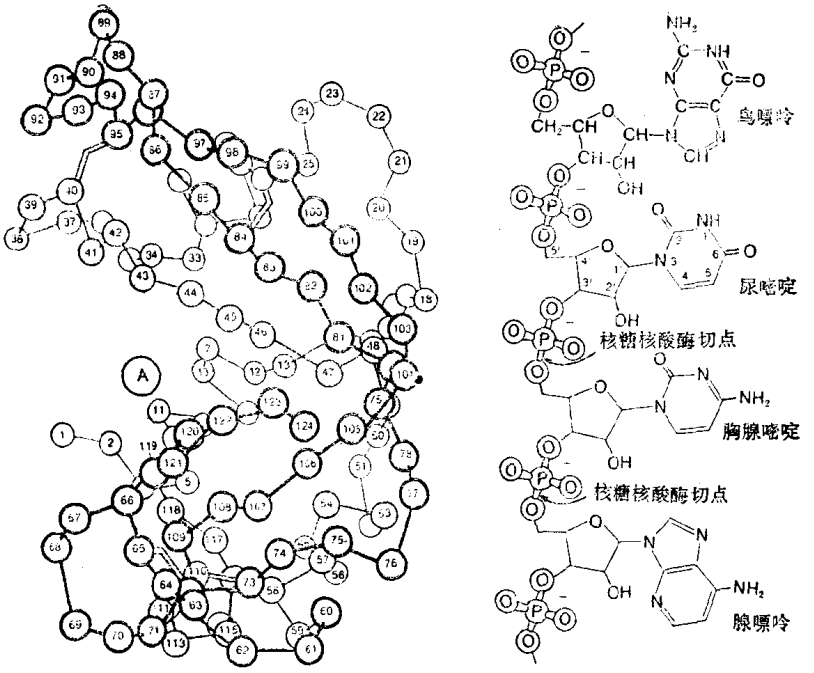


图 173　溶菌酶切割底物分子的假想机制

设想有二个氨基酸残基即Glu 35 和Asp 52 的侧链参与了溶菌酶切割底物分子的作用。反应机制包括由Glu 35 的羟基提供一个质子(H^+)结合到底物分子的D环和E环之间的氧原子上(见图 172)，从而切断底物D环和E环之间的糖苷键。由于D环上第 1 位碳原子失去氧原子带上一个正电(形成正碳离子)，处于D环邻近的 Asp52 残基侧链携带负电($—COO^-$)，它便吸引正碳离子暂时稳定化。此时加上一个水分子，其中 OH^-同正离子结合，H^+则取代Asp35 上失去的H^+，底物分子即一举而被切断。(引自 Dickerson 与 Geis 1969。)



(a) 核糖核酸酶A　　(b) 核糖核酸链(RNA)

图 174　核糖核酸酶A的三级结构

(a) 核糖核酸酶的活性部位槽用A标示。双线表示二硫键。
(b) 核糖核酸酶A的催化作用是在连接嘧啶残基的位置上切断RNA链。(引自 Dickerson 与 Geis, 1969。)

生物体系内往往有成千上万的化学反应同时进行着催化。个体发育过程中，新的酶和新的反应不断发生，旧的酶和旧的反应不断消失。整个系统的复杂性令人叹为观止，而随着对酶作用调节机理的不断揭示，其奥秘也逐渐大白于世。调节机理的研究进一步加深了人们对酶作用的认识，从而为研究生物体内代谢调控开拓了新的诱人前景。

酶活力的调节

调节是控制的一种形式，是为着管制某种能力以达到最适状态的一种手段。就最适宜的生长、维持、发育、行为、繁殖等所需的

条件来说，从未见到过这种情况：即要求所有的基因以尽其可能的速率制造转录本，而所有的转录本则又以尽其可能的速率翻译成蛋白质，乃至所有的酶以尽其所能的速率进行催化作用。促使转录、翻译和催化过程达到一般最适速率的动力称为调节作用力。现在看来，似乎每一种在一定方向上加强代谢作用和生理反应的动力，都存在着一个与之相对应的反作用力。调节机制能够保持这些不同的作用力达到平衡协调，平衡点一般受生物学上的最佳状态限定。

调节机制广泛存在于整个生物界，数量惊人，名目繁多，目前所知者不过十得其一。其范围从调节单个的酶作用开始，直到调节一个内部平衡的生态系统，或是一个开放平衡的生态系统。本节拟只讨论其中几例。

通过反馈抑制控制酶的催化是其中最简单的调节方式之一。前面谈及的关于活性部位和整个酶分子大小相差悬殊的问题，可以从二方面给以解释：(1)整个分子，或者至少其中某些部分起有定位和"操纵"活性部位的作用；(2)酶分子具有的多肽链类型（系指组成酶的亚单位——译注）常常在一种以上，而每类多肽链各有一个表面活性(或结合)部位。有一类表面活性部位称为调节部位。反馈抑制即可由这种表面活性部位得到解释。

反馈抑制是属于简单反馈环的一个典型例子（图 175)。由反应系列的末端产物抑制催化第一个反应的酶活力，亦即终止制造末端产物的信号系由末端产物本身提供。一当末端产物作为一种能源或者作为一种构造元件消耗殆尽，受调节的酶分子(调节酶)即被解除抑制又得以重新进行原先的合成反应。

大肠杆菌中天冬氨酸转氨甲酰酶(ATC酶)是导致嘧啶合成的代谢途径中催化第一个反应的酶，也是关于调节酶中研究调节机理的一个极好的例证。ATC 酶的催化活力可被（不是全部）嘧啶代谢(图 176)的一个末端产物胞苷三磷酸(CTP)所抑制。

ATC 酶已从提取较纯的细胞成分中分离获得，曾就其物理性质作了研究。全酶的沉降系数为 11.8S(接近于 300,000 道尔顿)。

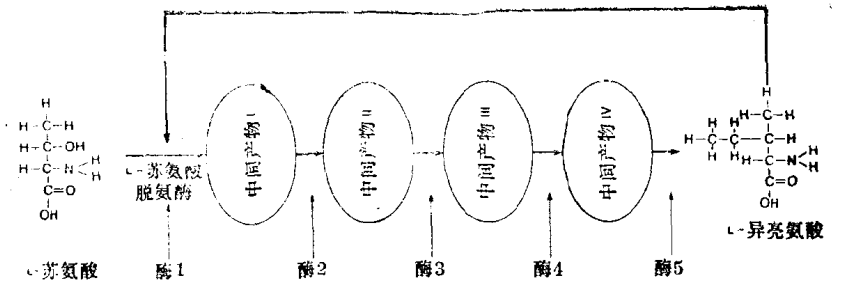


图 175　反馈抑制

关于L-异亮氨酸如何通过抑制合成途径中的苏氨酸脱氢酶(酶1)来调节自身合成速率的情况，可以从上述略图中得到说明。在这里酶1的活力是受细胞内异亮氨酸浓度调节的，而后者则反映了合成和消耗的相对速度。异亮氨酸同苏氨酸脱氢酶上特殊部位结合后，便改变了酶的构象(四级结构)，从而使之不能再去识别底物分子。(引自 R. P. Wagner, 1964。)

这个11.8 S的酶分子又能分成5.8 S和 2.8 S二个组分(亚单位)(图177)。在 5.8 S 亚单位中保留有催化活力；2.8 S 亚单位则与CTP结合。当 5.8 S亚单位同 2.8 S组分彼此分开时，CTP 不能抑制催化作用。但是当二个组分重新组合为一时，催化活性又对 CTP 的抑制表现敏感。

具有催化活力的 5.8 S组分是由三个 35,000 道尔顿的单体聚合而成的三聚体，其中每个单体均有一个与底物结合的部位(活性

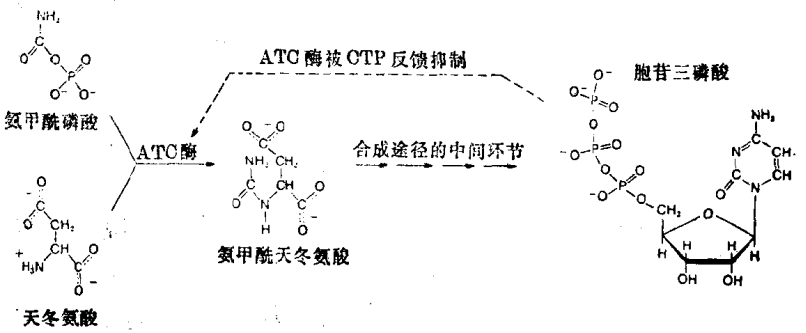


图 176　天冬氨酸转氨甲酰酶的调控机理

在胞苷三磷酸(CTP)合成、代谢途径的一系列反应中，其第一个反应由天冬氨酸转氨甲酰酶(ATC酶)催化。而ATC酶本身又能被CTP所抑制，即通过反馈抑制的方式来调节CTP的产量。

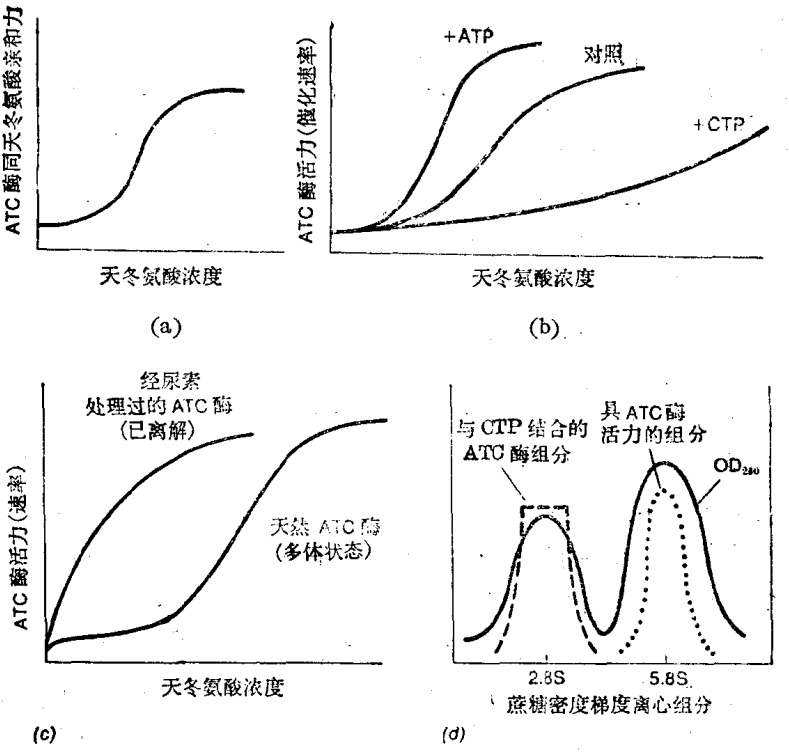


图 177　天冬氨酸转氨甲酰酶(ATC酶)的物理性质

(a) 天冬氨酸(Asp)为ATC酶底物之一。随着天冬氨酸浓度增加，ATC 酶活力亦表现相应增加。参见图中S状曲线所示。

(b) ATC 酶催化速率亦与亲和力一样。随天冬氨酸浓度的增加而递增。如图所示，CTP 能抑制ATC酶的催化活力，但不能完全抑制。增加ATP则能提高反应能力。

(c) 尿素处理能使ATC酶对CTP的抑制作用不再敏感，从而增加了同天冬氨酸的亲和力。这是因为尿素可将 ATC酶的多聚体离解成较小的聚合体，使其丧失亚基之间的活力协作性，图示即为由动力学曲线的S状变为双曲线状。从亚基离解所招致的动力学变化可知四级结构的重要性。

(d) 由蔗糖密度梯度离心测知，11.8S的 ATC 酶离解后产生2个较小的组分(一个亚基为5.8S，一个亚基为2.8S)。通过光密度 (OD_{280}) 图象比较 CTP 存在情况下的二个组分的酶活力。结果发现，5.8S 有催化活力，对CTP的抑制作用不敏感，2.8S组分同CTP结合，但无催化活力。(引自 J. Gerhart, 1964。)

部位)。另外与CTP结合的2.8 S组分也包含由二个17,000道尔顿的单体聚合而成的二聚体，其中每个单体各同一个CTP分子相

结合。因而完整的 ATC 酶是一个异源多聚体。ATC 酶大约 65% 是 5.8S 组分，约 35% 为 2.8S 组分。由这一百分比再根据二类单体的道尔顿数估算，作为寡聚体的该酶的结构组分应为：$3R_2+2C_3=R_6C_6$。这里 R 代表一个调节单体，C 代表一个催化单体，R_6C_6是天然酶(寡聚体) 的组合方式。在天然 ATC 酶中，抑制物和底物各自同酶结合的部位互不重叠，即二者同酶的结合位点位于蛋白质构型中彼此隔开的区域。由 CTP 进行的调节，确切些说，并非是因 CTP 直接与活性部位结合的结果，而是信号经 R 多肽传送到与之联系的 C 亚单位，最后到达活性部位。亦即，CTP的调节作用是以 ATC 酶的四级结构为媒介的。

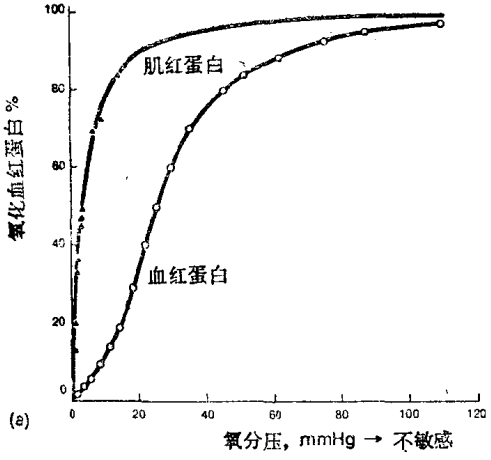
同溶菌酶和肌红蛋白相比，ATC酶属于一种大型的蛋白质，催化功能却并不过分复杂。据目前所知，它的大小和功能同它作为一个调节酶的角色有关，一般由酶分子中无识别催化底物能力的 R 单体来执行其控制调节。

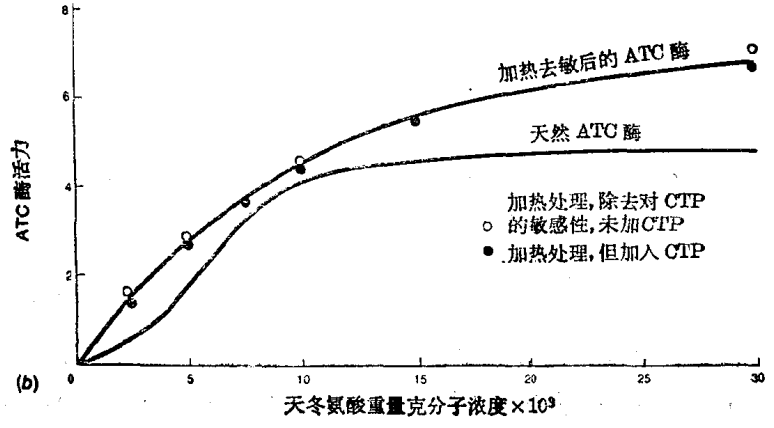
一个多聚体酶的四级结构往往会给反应速率以深刻影响，这种作用植根于多聚体酶的四级结构所具有的变构行为。下面试就变构行为问题，略述一例：A 为多聚体酶，假设是一个同源四聚体，具有同时识别至少四个底物分子的能力，每个单体各识别一个底物。酶对第一个底物分子的亲和力较低，当同第一个底物分子结合之后，酶对第二个底物分子的亲和力即有所加强。依此类推，同第三个底物分子的亲和力更强，同第四个底物分子的亲和力最强。比如血红蛋白具有四个同氧结合的部位，它同最后一个氧的亲和力约为第一个的 280 倍。这便是变构行为的结果。ATC酶也同样有变构现象 (图 178)。对于这种特殊的变构行为，即一个蛋白质对一个小分子结合作出的反应，一般即称为位置协同性。

多聚体酶的变构行为能以四级结构的构象渐变给以解释。在第一个结合反应发生之前，酶分子在总的外形上不易使活性部位同底物分子接近。经过第一个结合反应，诱导出某种构象变化，这个变化由蛋白质的结构传输到其余三个活性部位，使它们的反应能力得到加强。第二个结合反应同样能通过构象变化使后面二个

结合部位的亲和力进一步强化。最后，由于第三个结合反应的完成，第四个活性部位同底物分子的接近便再也不存在什么障碍了。四级结构在变构作用中的重要性从下述事实来看尤其明显：在ATC 酶中C 亚单位的变构行为系依赖于R亚单位的存在，即依赖于酶分子天然的四级结构。如果酶分子仅含有C 亚单位，虽然它仍表现有催化活力，但此时不再会有保证协同作用的配位基结合，当然亦不存在CTP 调节的问题。

多聚体酶还可通过反馈抑制之外的各种方式达到催化活力的





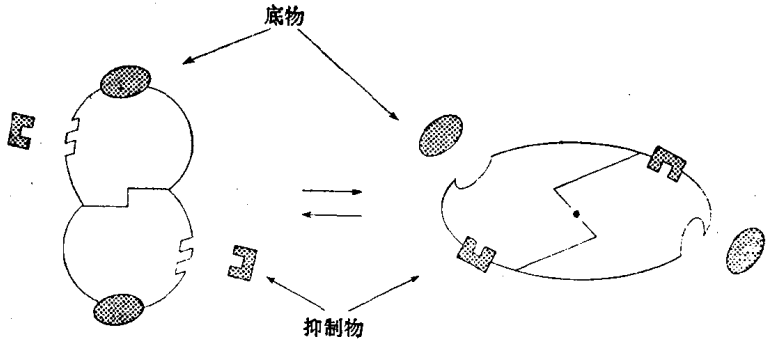


图 178　多聚体酶的协同作用及其变构行为

(a) 这是肌红蛋白和血红蛋白与氧结合的速率比较。每一个肌红蛋白分子结合一个氧分子，而每一个血红蛋白分子结合四个氧分子。图中肌红蛋白曲线为一双曲线，血红蛋白则为S形曲线。S形曲线充分反映了四级结构的协同作用，即随着与底物分子结合的亚基数增加蛋白质对底物的亲和力亦相应提高。在这种情况下，第一个氧分子的结合增加了蛋白质对第二个氧分子的亲和力，第二个氧分子结合又更增加了对第三个的亲和力，依此递增，直至血红蛋白分子同氧分子的结合完全达到饱和为止。

(b) 这是天然ATC酶的催化速率同加热去敏的 ATC 酶的催化速率比较。图中天然蛋白质的S形曲线是变构行为或者协同作用的结果。热去敏的ATC酶的双曲线则与 5.8S 亚基的情况相同，显示出无CTP的抑制作用，亦即不存在协作结合(图 177c)。(引自 Gerhart, 1964,及 J. P. Changeux, 1965。)

(c) 为了在分子水平上解释变构行为，这里设想一个双聚体酶模型，酶分子的每个单体有三个结合部位。本图可以模式地说明，由底物诱导所产生的构象变化导致了对抑制物不敏感反应，而由抑制物所产生的构象变化导致了对底物不敏感反应。此外，在某些酶中还存在有一个激活剂的结合部位。所谓激活剂乃是因为与酶结合后，增强了此酶对底物的亲和力。(引自 Changeux, 1965。)

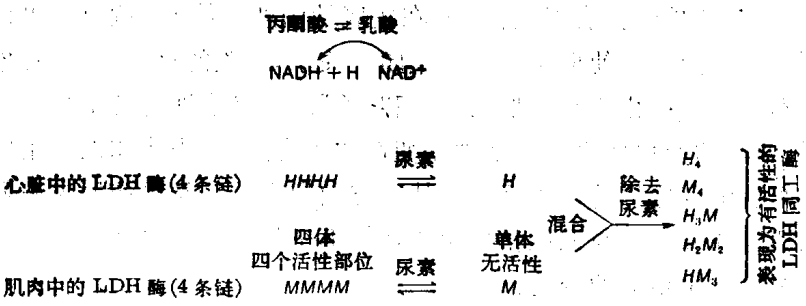
调节。例如，对大多数酶来说，在煮沸的条件下一般均能使之变性和受到损伤。可是在细菌中，有一种会分解底物烟酰胺腺嘌呤二核苷酸 (NAD) 的酶却需要预先煮沸才表现有催化作用。这种天然酶的复合体兼具有热稳定和热不稳定两种功能。该酶大小约为52,000个道尔顿，调节亚单位约 35,000 道尔顿，催化亚单位约17,500道尔顿。目前对这种匪夷所思的反常调节方式尚无从得到解释，有人推测由于个体代谢的需要，在与之相应的较高温度条件

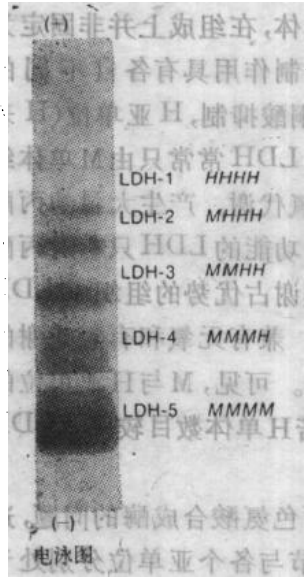
下，调节亚单位会表现钝化，而在较底的温度条件下，调节亚单位则行使功能——使催化亚单位关闭。关于开-关的这一调节现象并不是二种静止状态，如同"开"和"关"二个字表面含意那样；相反，速率的变化倒更象是某种渐变性的动态进程：当诸如温度条件或产物浓度这样的环境因素向某一方向递变时，结果使原先有催化功能的酶分子与无催化功能或功能不全的酶分子之间数量上的比例也发生相应的变动。

另一个关于四级结构控制调节的例子是乳酸脱氢酶（LDH）（图 179）。LDH 含有二种类型的四个单体，在组成上并非固定为 1:1。这二种类型的亚单位对丙酮酸的抑制作用具有各自不同的敏感性。M亚单位（M 表示肌肉）不受丙酮酸抑制，H 亚单位（H 表示心脏）则受丙酮酸抑制。肌肉组织中的LDH 常常只由M单体组成，这是因为肌肉组织中经常进行着无氧代谢，产生大量的丙酮酸。因此，在这样一种特定的环境中行使功能的LDH只有对丙酮酸不敏感才能适应。而处于心脏和有氧代谢占优势的组织中 LDH 便由H单体组成，并对丙酮酸抑制敏感。兼有无氧和有氧代谢的组织则常有按 2H:2M 组成的 LDH 分子。可见，M与H亚单位的比例一般由氧呼吸状态来调节。同样，若H单体数目较多，LDH 活力受丙酮酸抑制的程度也愈高。

在遗传编码一章中，曾提到大肠杆菌色氨酸合成酶的问题，这是有关酶调节机理的又一例证。这种调节与各个亚单位分别处于组装态和解离态时在催化性质与产物产量方面的差异有关（表 25）。在这里，α 亚单位有促进反应Ⅲ和抑制反应Ⅳ的作用（变性的 α 亚单位对上述二个反应均无效应）。

目前尚无法得知在 α 亚单位和 β 亚单位之间所观察到的这种关系是否真实反映了体内控制的某种机制。为了能言之有据地从体外系统推知体内条件下的某些情况，还必须要做许多补充实验（包括体外实验中单个酶的作用与体内多个不同酶相互作用的差别）。例如，在一定的代谢途径中，所有中间产物的浓度均取决于每个反应的平衡常数（按质量作用定律由反应速度系数计算）。有





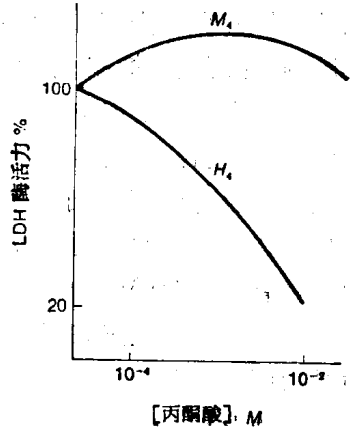


图 179 乳酸脱氢酶(LDH)

LDH在反应中催化丙酮酸转变成乳酸。这种酶由四个亚单位组成，哺乳动物中有二种亚单位，一种在心脏中占优势(H)，另一种在肌肉组织占优势(M)。在其他组织中H:M的组成比率随细胞中的丙酮酸浓度而有变化。H亚单位对丙酮酸敏感，M亚单位对丙酮酸不敏感。心脏中LDH酶的亚单位组成为HHHH(与心脏中丙酮酸浓度很低有关)，肌肉中 LDH 酶为MMMM(与肌肉中丙酮酸浓度较高有关)。LDH酶共有五种结构类型(4H，3H-1M，2H-2M，1H-3M和4M)，又称为同工酶。因为M和H亚单位具有不同的电荷，故在电场中可将五种类型的同工酶相互分离出来。其中异源同工酶（含H-M 二种亚单位)的电泳迁移率分别处在同源同工酶MMMM和 HHHH 之间。二种亚单位对丙酮酸浓度增加的反应互不相同（见图)。（同工酶电泳图引自 Markert 与 Apella，1963。)

表 25　色氨酸合成酶亚单位成分与催化活力特性、活力强度之间的关系

反　应	底　　物	产　　物	α	β	$\alpha_2\beta_2$*
I	吲哚磷酸甘油＋丝氨酸	色氨酸	−	−	＋
II	吲哚磷酸甘油	吲　哚	＋	−	＋
III	吲哚＋丝氨酸	色氨酸	−	＋	++
IV	丝氨酸	丙酮酸＋NH_3	−	＋	−

* $\alpha_2\beta_2$ 是色氨酸合成酶具有正常生理功能的组成状态，反应 I 是生理上有重要意义的一步。

＋＝多肽链(单体或多聚体)具有催化活力。

－＝无催化活力。

++＝催化活力更高。

关的酶促反应在上述动态平衡中理论上都是可逆的，但实际上处于代谢恒定状态中的体内反应基本上只能朝一个方向进行（因为代谢产物在体内不断消耗和分配，使平衡向产物生成方向移动——译注），这使得在确认各个酶在整个恒定系统中所起的具体效应方面发生了困难。当然，除非此种效应具有某种极端性质，诸

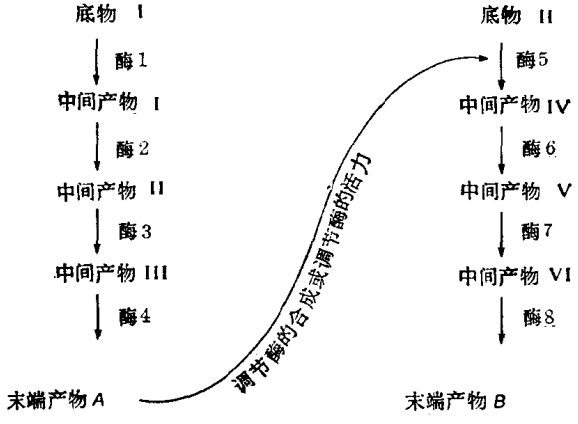


图 180　代谢途径之间的相互调节

一条代谢途径的末端产物可以通过控制酶合成或控制酶活力等方法调节另一条代谢途径中的酶的催化作用。图示以交互作用方式调节代谢的二组酶促反应系统。

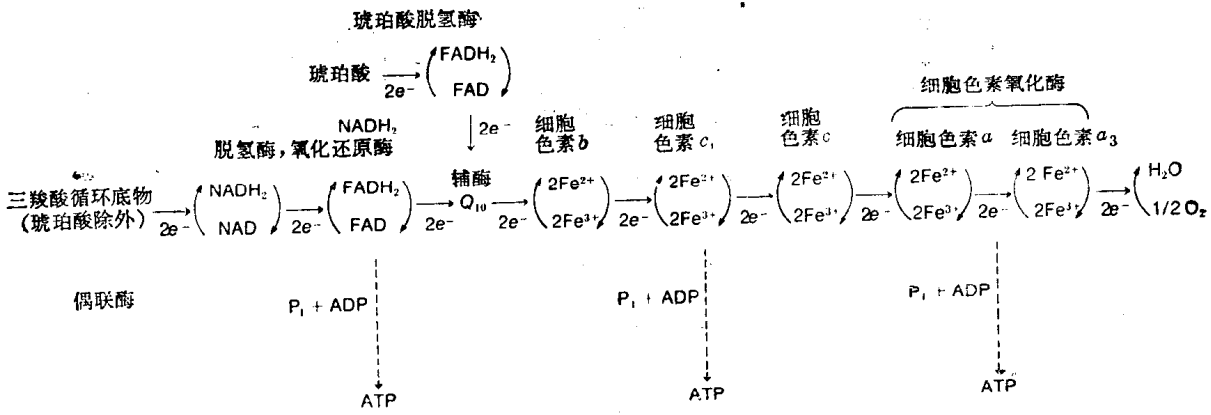


图 181 偶联反应

偶联反应的速率受到一组酶系的影响。例如，电子从左向右传递依靠电子供体(失去电子变成氧化态的细胞色素)和电子受体(接受电子变成还原态的细胞色素)之间的接力式反应。流经细胞色素的电子流可控制与之偶联产生ATP 的磷酸化反应，因而赋有如下含义：即这种调节方式可视为是统一体系中的许多相互竞争又相互依赖的组分之间的某种动态平衡。

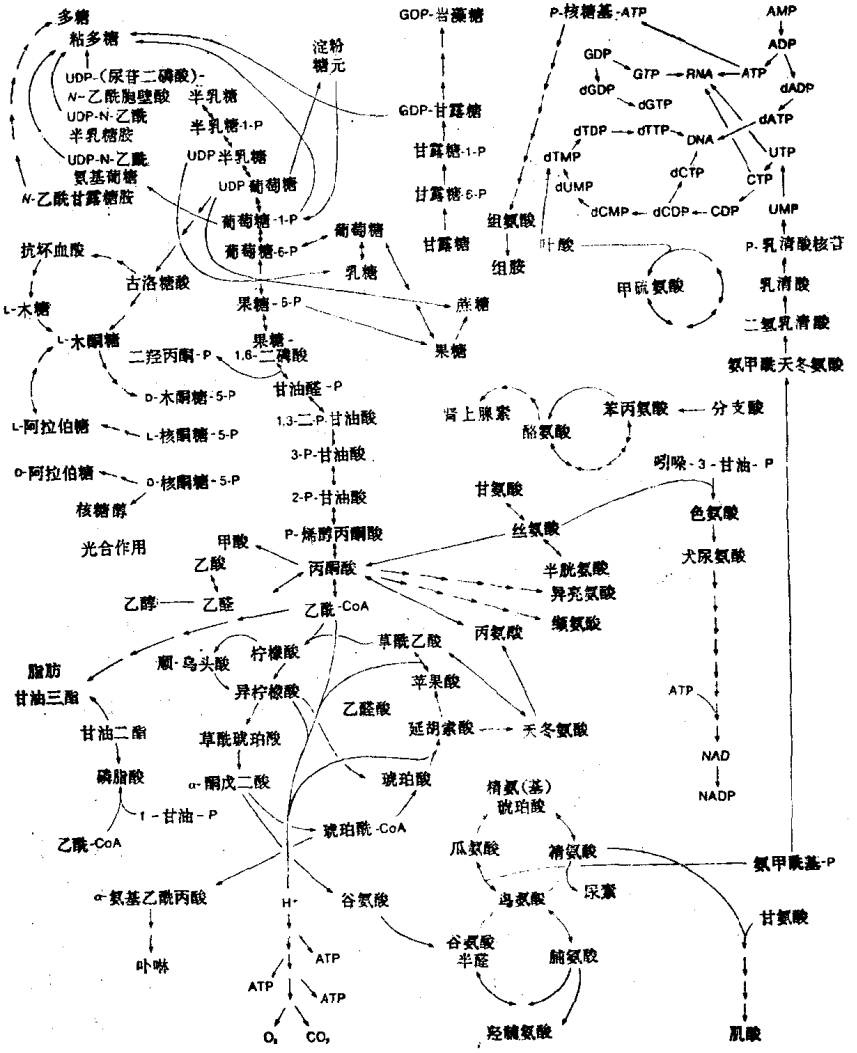


图 182　代谢途径中的锁连关系

看一眼这张迷宫般的代谢路线的锁连图便会知道，在合成代谢产物，将代谢产物配置到起作用的位置以及在调节代谢产物的浓度方面，必定存在有各种不同的方式。从基因控制代谢，有关基因调控机理，这二者进化历史的角度来看，代谢锁连关系中的瓶颈效应并未减弱。

如象阻断代谢径路进程，或者极大地增加反应速率等等，才能从中间产物浓度的变化上，对有关某个酶加以检测。

如果某些代谢路线由于共同的底物或共同的末端产物相互受到牵制，或者某一代谢途径的末端产物成为另一代谢途径的底物的抑制物时，调节系统便会显得极其复杂（图 180）。如象在氧化磷酸化反应的电子传递系统中（图 181），则由专门的偶联反应将各个酶串联在一起，整个地作为一个单元进行调节。在电子传递系统中，酶组（系）的活力是同耗氧量的高低以及与 ATP 的产量（由ADT，无机磷酸产生）相对应的。ATP 的产率本身则又受到细胞内其他生理过程的调节。在这里，所有由酶催化的反应都具有偶联性质，因此，没有一个反应能够脱离其他反应而独立存在。

关于各条代谢途径之间的相互关系，图 182 可以给我们若干启示。人们可以看到，在许多代谢途径中酶的活力不高不低，往往恰到好处。事实上，即使识别同一个底物的酶之间亲和力相同，但由于复杂代谢路线中的某些瓶颈和堵塞，结果仍能造成这些酶之间的活力差异。这里只要从食物消化的角度来看，就不难理解这一简单的关系了。人体摄取碳水化合物、脂肪、蛋白质等物质作为营养，其中每一种成分都能被氧化并生成 ATP、GTP 等用以提供能量。但是，人体利用这些物质的能力却受如下方式的调节：当所有这三种物质同时存在时，作为能量利用的首先是碳水化合物，要等碳水化合物耗尽之后，方才启用脂肪，而蛋白质的分解则要排到最后一个。由此可见，这类调节方式是非常普遍的。但是，除开酶的不同亲和力所发挥的作用之外，有关操纵调节的详细机制还不得而知。在原核生物中，另外还有证据表明，行使调节作用还可以通过每个细胞中酶分子的数目来完成，而酶分子的数目则又受控于蛋白质的合成。

酶合成的调节

除了通过酶活力的反馈控制调节代谢需要外，细胞的另一种调节形式是控制酶的产生。蛋白质的合成机制暗示，在控制酶浓

度方面，这里可能有几种方式：例如控制核糖体和 mRNA 分子的结合，控制mRNA分子的周转速率以及调节转录和(或)转译速率等等。看来在细菌系统中，控制mRNA的合成似乎比其他的方式更为重要。

就一般情形而论，调节细胞内酶分子的数量可从二方面进行：(1)诱导合成细胞中不存在的酶，(2)抑制细胞中现有酶的再形成。诱导表示酶合成的启动，抑制则标志着酶合成关闭。上述二过程在细菌系统中已经都有明例。

在大肠杆菌中，了解得最清楚的诱导系统之一为β-半乳糖苷酶(β-gal)，该酶能将乳糖分解为葡萄糖和半乳糖二种单糖(图183)。

生长在以葡萄糖为唯一碳源上的大肠杆菌，平均每个细胞约含有一个 β-半乳糖苷酶分子。(每一细胞的 β-半乳糖苷酶分子数目可借助于一种带有人造萤光物质的底物与酶结合后，使用萤光计测定，其敏感程度甚至连单个酶分子亦可检测。)可是一当细菌转移到缺乏葡萄糖但含有乳糖的培养基上生长之后，不过二分钟左右，细胞即开始合成 β-半乳糖苷酶。最后，每个细胞大约能含有 3,000 个酶分子。此时，与 β-半乳糖苷酶一起被诱导的还有

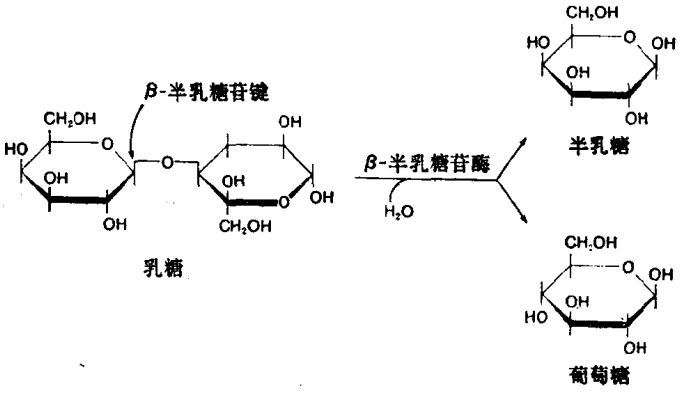


图 183　β-半乳糖苷酶

β-半乳糖苷酶催化双糖(乳糖)分解为单糖(半乳糖和葡萄糖)的水解反应。不能制造具有功能的β-半乳糖苷酶的突变型便不能利用乳糖作为碳源。

另外二种酶。一种为乙酰基转移酶，功能不明。一种为透性酶，能使乳糖分子更易进入细菌细胞。许多在化学组成与结构上与乳糖类似的化合物一般都能诱导这三种酶合成(表 26)。

表 26　各种半乳糖苷诱导 β-半乳糖苷酶和半乳糖苷转乙酰酶合成的情况

化　合　物	浓度(M)	β-半乳糖苷酶	半乳糖苷转乙酰酶
β-D-硫代半乳糖苷:			
异丙基	10^{-4}	++	++
甲　基	10^{-4}	+	+
	10^{-5}	±	−
苯　基	10^{-3}	−	−
苯乙基	10^{-3}	±	±
β-D-半乳糖苷:			
乳　糖	10^{-3}	+	+
苯　基	10^{-3}	±	±
α-D-半乳糖苷:			
密二糖	10^{-3}	+	+
β-D-葡糖苷:			
苯　基	10^{-3}	−	−
半乳糖	10^{-3}	−	−

++＝高效诱导，+＝良好诱导，±＝弱诱导，−＝无诱导。

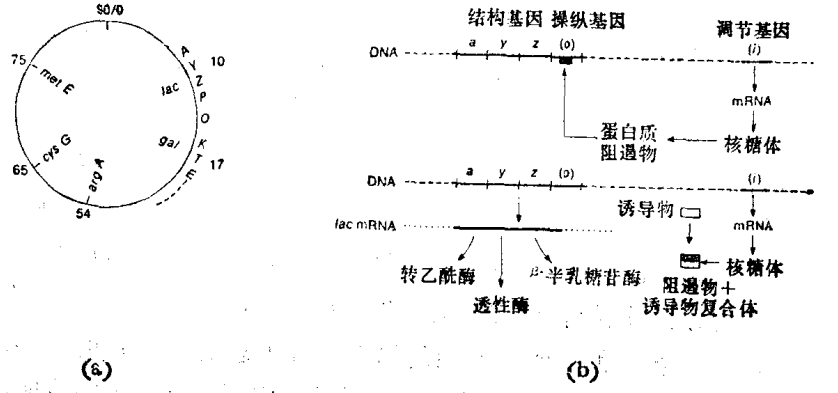
这三种酶(β-半乳糖苷酶，乙酰基转移酶和透性酶)均同时受一种诱导物，即乳糖的控制，可以这样解释，即编码这三种酶的三个结构基因在细菌染色体上彼此相连，并且被转录到同一个多顺反子 mRNA 转录本上。转录从基因顺序的β-半乳糖苷酶一端起始，每次制造一个β-半乳糖苷酶基因转录本。当然在这个 mRNA 转录本上，同时也包括着其他二个基因的转录本。

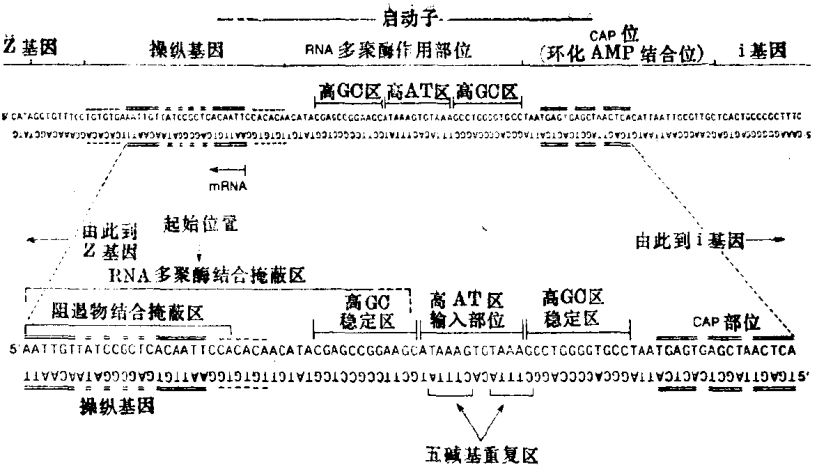
染色体上以一个或多个结构基因为单位调节 mRNA 合成的一段区域称为操纵子。这个用语是由因揭示了乳糖操纵子 (*lac*)

详细情节而举世闻名的 Jacob 和 Monod 选取的。*lac* 操纵子中(图 184)除了三种结构基因之外，还包括染色体的其他区域。简单说来，*lac* 操纵子中的三个结构基因是 *z*、*y* 和 *a*。半乳糖苷酶、透性酶和乙酰转移酶的 mRNA 转录本即分别以它们为模板而产生。*lac* 操纵子上的操纵基因区域(*o*)可识别由调节基因(*i*)编码合成的阻遏蛋白质。此外 *lac* 操纵子上还有一个启动基因 (*p*)。RNA 多聚酶在转录结构基因之前，必须要先与*p*结合。有关操纵子的这些成员，下面还要详细讨论。

对于 *lac* 操纵子上的 *i* (一般认为调节基因 *i* 不属 *lac* 操纵子区域——译注)、*o*、*p* 各个基因可以通过分离获得各个基因位置上的不同突变型加以识别。例如有一类突变型不能在结构基因占据的染色体上恰当地作图定位。但可由这三种酶在突变细胞中与正常型相比的相对浓度变异而鉴别 (图 184)。这些突变型以存在诱导物时的特有速率(称组成速率)来合成这三种酶，即使诱导物耗尽也不停止，似乎它们是处于正常的控制机制之外。已知发生这些突变的基因叫做 *i* 基因，因而提出 *i* 基因的作用就在于调节结构基因的转录。

为了有助于认识 *i* 突变型如何表现其特性，以及对 *i* 基因是怎样确定其调节功能的问题，有必要先介绍一下在大肠杆菌系统中采用的遗传学实验技术。人们可以操作细菌细胞，以使其基因





(c)

图 184 *lac* 操纵子

(a) 图中所示环状基因图系按遗传物质的转移时间绘制的细菌染色体图。染色体转移在*o*处开始，按顺时针方向进入受体细胞。按时间间隔中断杂交并检查受体细胞的表型。结果表明，在染色体开始转移 10～12 分钟之间，*lac* 操纵子即进入受体细胞。整个染色体从供体细胞移入受体细胞共需 90 分钟左右。

(b) *lac* 操纵子中许多突变部位已通过经典遗传学的实验方法加以鉴别。图示二种情况中的 *lac* 操纵子，一为关闭状态(上行)，此时不合成β-半乳糖苷酶；一为开放状态(下行)，此时合成β-半乳糖苷酶。在关闭状态，由*i*基因编码的阻遏物同操纵子的操纵基因 (*o*) 相结合，以阻止结构基因的转录。而在开放时，阻遏物同诱导物分子(乳糖)先形成复合物，亦即诱导物分子能抑制阻遏物再同*o*基因结合。一旦阻遏物离开*o*基因，结构基因便立即开始转录。

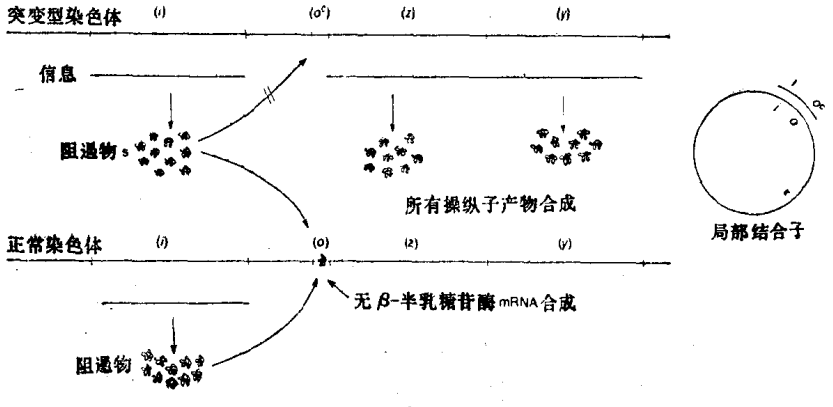
(c) 图示*i*基因和*z*基因之间核苷酸的排列顺序。用横线标志者为高GC区和高AT区，各碱基配对区。阻遏物和 RNA 多聚酶同基因结合的部位亦处于这一区域。(引自 W. Barnes, R. Dickson, B. Reznikoff, 与 J. Abelson, 1975。)

组的部分基因成为二倍体。这类细胞即为部分双倍体，或称局部接合子 (图 185)。局部接合子具有一条完整的染色体组，再加上一段同源的染色体片段。位于"双倍化"区域里的基因可以进行显隐性遗传分析，就如在正常双倍体生物中所做的那样。

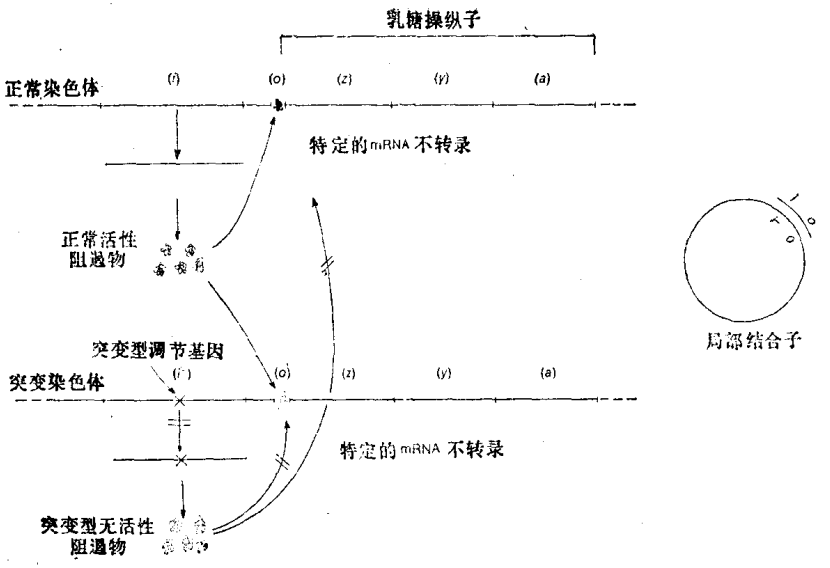
含有一个 i^-和一个 i^+基因的局部接合子总是呈现 i^+的表型，这意味着 i^-/i^+双倍体是诱导型，而不是组成型。这一事实表明，

位于附加染色体片段上的 i^+基因产生一种物质，此物质可通过细胞质输送到同源的、完整的染色体上某一部位，并在该部位抑制转录。这一观察所用的细菌类型为基因型 i^-z^+/i^+z^-的局部接合子，因此，这样一个品系在诱导物缺少时不能合成 β 半乳糖苷酶。这也意味着附加染色体片段上的i^+基因产生的调节物质抑制了位于不同染色体上 z^+基因的转录。如果此时 z^+ 基因为各种 z 突变所取代了的话，那么其中每一种 z 突变都可编码一种变异了的 β 半乳糖苷酶。每一种z 突变都以同一方式准确地对 i^+基因产生的调节物质作出响应，其结果是每一种突变了的酶都可由乳糖诱导。

几乎在发现 *lac* 操纵子被一种称为阻遏物的调节物质调节的同时，分离和鉴别阻遏物的工作也即开始。最初，人们一度相信阻遏物分子是RNA，以为 RNA 是基因产物又能与 DNA 互补，因而亦能与操纵子DNA结合从而中止转录。但 RNA 在体内产生后很快降解，不能长久保持随时备用。后来又发现阻遏物的活力可被蛋白质合成抑制剂所抑制，i 基因的某些突变对温度表现敏感，以及由于将中止突变(UAG 密码子)引入 i 基因而能中止阻遏物产生，等等。(不妨回顾一下，UAG 密码子只可以终止蛋白质合成而不能终止 RNA 合成)。于是，蛋白质便理所当然地成了人们注意的对象。



(a)



(b)

图 185　大肠杆菌部分二倍体(局部接合子)

一个局部接合子细胞含有一条完整的染色体和另一条附加同源染色体片段。采用局部接合子，使染色体 *lac* 操纵子区域二倍化，即可研究调节基因突变型之间的显隐性关系。比如可以得到 i^+和 i^-及(或) z^+和 z^-的组合(基因对)。假若 i^+, z^-基因位于附加染色体片段上，而 i^-, z^+位于完整染色体上，那么 i^+基因就会将细胞的表型从组成型变成为诱导型。

(a) 这是含有一个正常操纵基因(o)和一个突变操纵基因(o^c)的局部接合子。在局部接合子中o^c对o为显性，由于o^c不能同阻遏物结合，因而操纵子成功转录，产物便不断合成。

(b) 这是一个含有正常调节基因(i)和突变调节基因(i^-)的局部接合子。i基因对 i^-为显性，因为它能产生正常的具有活性的阻遏物。阻遏物可同时作用于二条染色体上的操纵基因，故操纵子中所有结构基因的转录均被抑制。(引自 Watson, 1970。)

Walter Gilbert 是分离阻遏物蛋白质成功的第一个学者。他运用的原理是基于这一假想：既然诱导物分子会抑制阻遏物分子的抑制转录作用(图 186)，那么阻遏物分子必定会优先同诱导物结合。在研究 i 基因的产物中，Gilbert 采用了一种不同于乳糖的乳糖同系物，即异丙基-β-D-硫代半乳糖苷(IPTG)，这是因为 β-半乳糖苷酶不能够降解这种同系物的化学结构。实验方法在技术性上实际相当简单：将细胞提取物加上放射性 IPTG 分子放入一

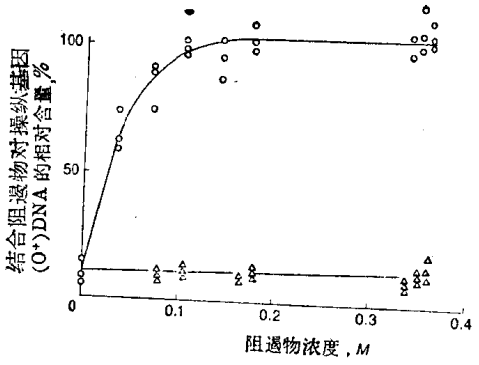


图 186 *lac* 阻遏物同 DNA 的结合

诱导物(乳糖)缺少时阻遏物同操纵基因DNA(○)结合;诱导物存在时,阻遏物不同DNA(△)结合。实验中 DNA 浓度始终保持恒定,改变的乃是阻遏物的浓度。

个半透性的透析袋中,然后将封好口的透析袋置于一个大的盛有缓冲液的器皿里进行透析。未结合的 IPTG 分子是一种小分子,可穿过透析袋的壁膜进入缓冲液,而 IPTG 与蛋白质形成的复合物由于分子量太大不能穿过透析袋膜而贮留在透析袋里。预先将 IPTG 分子作好放射性标记,即使阻遏物分子的浓度较低,作者亦同样能够发现放射性信号,证明其中有与放射性 IPTG 相结合的蛋白质。

然而,如果仅仅由于存在着与 IPTG 相结合的蛋白质就据以断定这种蛋白质必定是 *i* 基因的产物,理由未免不够充分。关于分离得到的蛋白质来源问题,曾从其他类型的实验中得到进一步的证据。其中之一便是将来自许多 *i* 突变型的蛋白质进行比较,已知这些突变型产生的阻遏物在某些特定的性质上各有不同,所以通过 IPTG 与蛋白质结合进行分离之后,便可比较这类蛋白质彼此之间的差异,再同突变型之间的变异一一对照。从一系列这样的比较,终于得出了阻遏物是 *i* 基因产物的确切证据。

lac 操纵子的阻遏物为一大小约 150,000 道尔顿的四聚体,其中每个单体均可同一个 IPTG 结合。已经证明阻遏物同 *lac* 操纵

子的操纵基因区域相结合（图 184），至于阻遏物单体表面上的哪一部位能够识别操纵基因DNA并与之结合，尚未成功鉴定。阻遏物一旦同DNA结合，操纵子的转录即便停止。而当阻遏物同诱导物相结合，使阻遏物从DNA上被释放之后，于此同时，操纵子的转录也就马上开始。

lac 操纵子受阻遏物控制的机理比较简单，它们属于负控制调控系统（即活性阻遏物同操纵子结合抑制转录）并且效率极高。一个150,000道尔顿的阻遏物分子即可抑制一个4×10^6道尔顿DNA片段的转录（阻遏物重量为其所抑制基因的百分之三）。但是，人们并不知道生物种属中广泛存在的这种控制系统的来龙去脉。迄今为止，在高等生物中亦还未发现完全类似这样的控制体制。

在 *lac* 操纵子中还分离出了另一些突变型。其中 i 突变型中有一种叫做超阻遏突变型 i^s，它所编码的阻遏物不能识别诱导物分子。道理很简单，即 i^s 突变破坏了阻遏物同诱导物的结合部位，而识别DNA并与之结合的部位却完整无恙。由于 i^s 突变型产生的阻遏物不能识别诱导物，所以 *lac* 操纵子也就不能解除抑制而一直处在关闭状态。换句话说，i^s 突变型不能够利用乳糖作为碳源，因为不能产生分解乳糖所必需的酶。

操纵基因突变型 o^c 同 i^- 突变型一样也属于组成型，即转录始终"开放"不停。但 o^c 突变型亦有不同于 i^- 突变型之处，即操纵基因突变型不受同源染色体上野生型等位基因（o^+）的影响。o^c 突变的效应仅受与之相邻的结构基因限制。一般认为，o^c 突变型之所以会表现转录始终"开放"，乃是因为阻遏物不能同突变型操纵基因相结合从而丧失有效控制转录的能力。o 基因、i 基因和结构基因都有不同的突变类型，相互间不同的组合方式可产生相应的不同表型。参看表 27 所列资料，读者也许会获得一个大致的印象。

虽然 *lac* 操纵子的第四片段即启动区域尚未进行充分的遗传学研究，但人们已经知道它是RNA多聚酶附着的区域。这一事实还是在研究了操纵子中其他基因的细节几年之后才被人们发现。

表 27 *lac* 操纵子中不同基因突变的表型

基因型	表型	产生的 β-半乳糖苷酶	
		加入诱导物	不加诱导物
i^-z^-	lac^-	−	−
i^-z^+	组成型	+	+
i^+z^+	诱导型	+	−
i^+z^-	lac^-	−	−
i^-z^+/i^-z^+	组成型	+	+
i^+z^-/i^-z^-	lac^-	+	−
i^-z^-/i^+z^-	lac^-	−	−
i^-z^-/i^+z^+	诱导型	+	−
i^+z^-/i^-z^+	诱导型	+	−
i^-z^+/i^+z^-	诱导型	+	−
$i^+o^+z^+$	诱导型	+	−
$i^+o^cz^+$	组成型	+	+
$i^-o^cz^+$	组成型	+	+
$i^+o^cz^-$	lac^-	−	−
$i^+o^+z^+/i^-o^+z^+$	诱导型	+	−
$i^+o^cz^+/i^-o^+z^-$	组成型	+	+
$i^-o^+z^-/i^-o^cz^+$	组成型	+	+
$i^+o^+z^+/i^-o^cz^-$	诱导型	+	−
$i^+o^+z^-/i^+o^cz^+$	组成型	+	+
$i^+o^cz^-/i^-o^+z^+$	诱导型	+	−

操纵子中是否还会发现其他控制成分，目前还只是一种猜测。不过，在任何情况下，关于 *lac* 操纵子的基本模型恐怕是不会有多大改动的。值得更加重视的倒是应设法查明 *lac* 操纵子模型究竟是一类稀有的调节机制呢，还是广泛存在于生物种属之中的普遍现象。

lac 操纵子结构基因的调控机构还可被另一操纵子的调控机构例如色氨酸操纵基因所占用而取代(图 187)。若将 *lac* 操纵子的 *i*、*o*、*p* 段切开，再把色氨酸操纵子(*trp*)接到 *lac* 操纵子的结构基因上，结果调节色氨酸生产的信号将同时能指导 β-半乳糖苷酶、透性酶、乙酰基转移酶的合成。

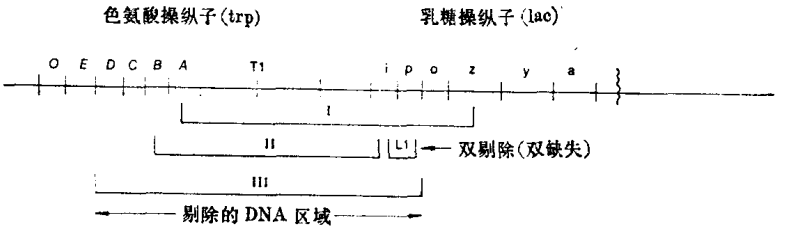
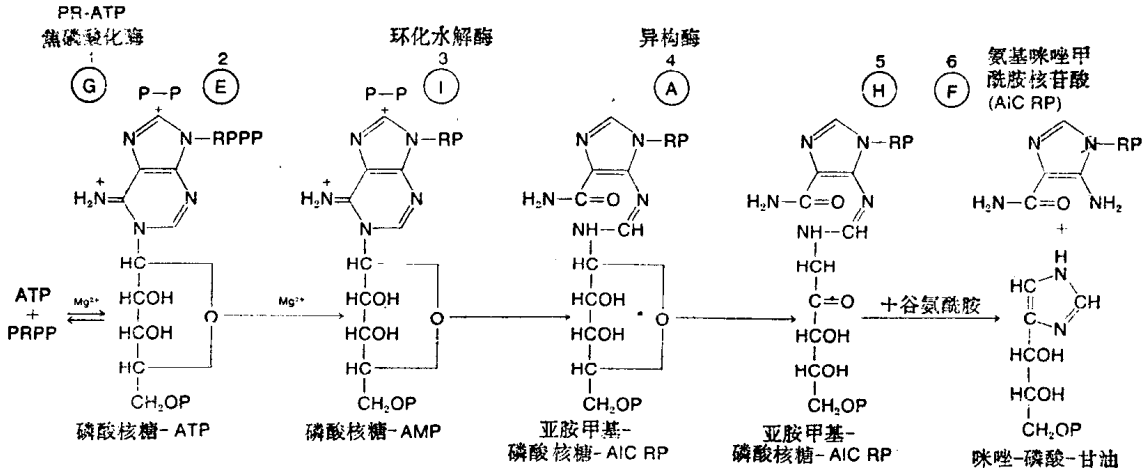


图 187　采用染色体剔除法使 *lac* 操纵子和 *trp* 操纵子在特定位点相互连接

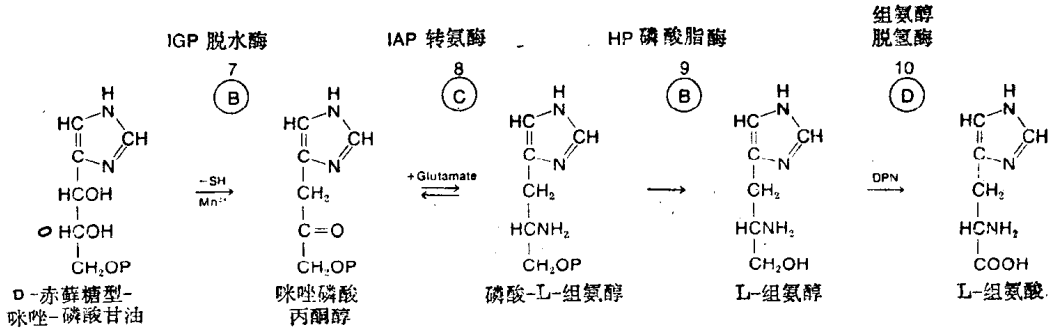
在 *trp* 操纵子和 *lac* 操纵子之间若将某段染色体剔除，然后再使断头接上，则 *trp* 操纵子的控制机构将有效地调节 *lac* 操纵子的结构基因。剔除区 I 涉及到 *lac* 操纵子的 *p*，*o* 基因及另一区段的 *i* 基因。对于这一型细胞来说，乳糖诱导物不诱导 β-半乳糖苷酶的形成。而色氨酸代谢系统的调控机制却可以控制 β-半乳糖苷酶的形成。

除了存在诱导型(即分解代谢调控)操纵子外，还有另一种阻遏型(即合成代谢调控)操纵子。阻遏型操纵子与诱导型操纵子大体类似，二者之间只有一个重大差别：在诱导系统中，诱导物分子可与阻遏转录的阻遏物结合以解除抑制；而在阻遏操纵子中，代谢途径的末端产物能使原先不起阻遏作用的阻遏蛋白活化从而行使抑制转录的功能。鼠伤寒沙门氏菌 (*Salmonella typhymurium*) 的组氨酸操纵子(*his*)是现今研究得最彻底的阻遏操纵子之一(图 188)。

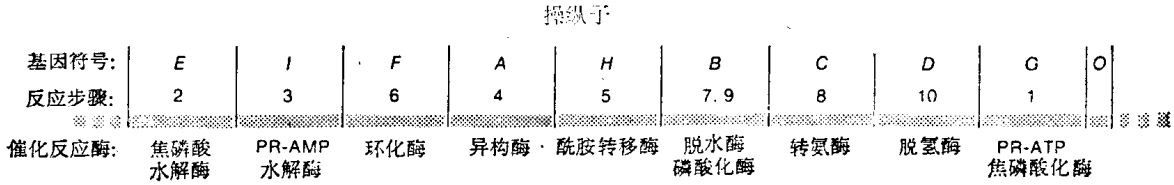
his 操纵子中有九个结构基因，它们共编码十个参与组氨酸生物合成的酶(脱水酶、磷酸化酶编码在一个基因 *B* 上——译注)。图 188 表示组氨酸合成的反应步骤、催化反应所需的酶和结构基因遗传学图之间的关系。操纵子的操纵基因同编码代谢中第一个酶的结构基因相邻，并位于所有结构基因的末端。操纵子中所有结构基因在转录时作为一个单位，因而产生一个共同的 mRNA，这


PR-ATP
焦磷酸化酶
1
G
2
E
环化水解酶
3
I
异构酶
4
A
5
H
6
F
氨基咪唑甲
酰胺核苷酸
(AIC RP)
ATP
+
PRPP
Mg2+
Mg2+
磷酸核糖-ATP
磷酸核糖-AMP
亚胺甲基-
磷酸核糖-AIC RP
亚胺甲基-
磷酸核糖-AIC RP
+谷氨酰胺
咪唑-磷酸-甘油

(a)(1)



(a)(2)



(b)

图 188 *his* 操纵子

(a) 这里表示的是体内合成组氨酸的反应步骤及其每步反应所需要的酶。

(b) *his*操纵子由九个紧密相连的结构基因和位于图中右端的一个操纵基因所组成。这些结构基因经转录所产生的多聚顺反子mRNA含有约 13,000 个核苷酸。信息的转录从操纵基因一端起始至E基因端结束，如中间有一个中止突变发生，转录即在此点中断。(引自 B. Ames, 1965。)

是迄今所研究的最大的多顺反子 mRNA 分子。抑制过程同转录相仿，所有结构基因都毫无例外地同时受到组氨酸产物的抑制。组氨酸在这里扮演的是一个辅阻遏物，在阻遏物和组氨酸结合之前，前者一直处于无活性状态。一旦二者形成复合物，便立即同操纵基因结合从而抑制转录。

关于组氨酸的作用方式，从分离得到的二种突变型研究中也已查明：是His-tRNA，而不是组氨酸本身激活了*his*操纵子阻遏物。第一，使*his*操纵子调节失灵的突变位于远离*his*操纵子的染色体区域，从而排除了操纵子上基因突变的可能性。第二，这些突变使

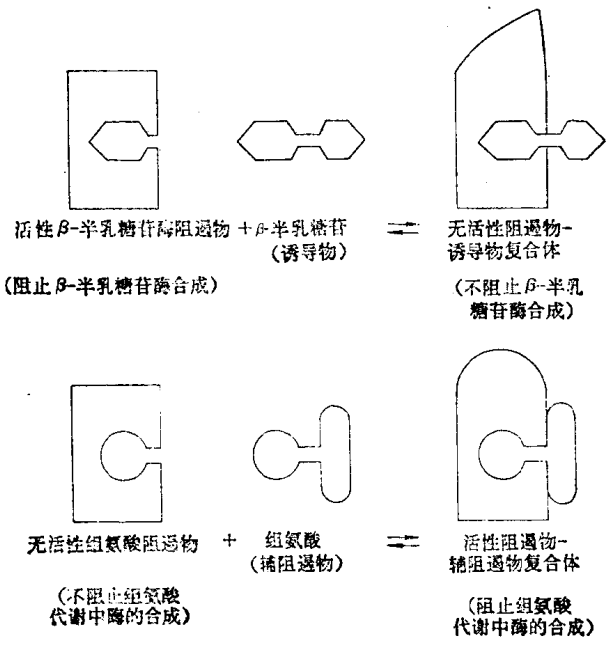


图 183 诱导模型和阻遏模型

这二个模型概括地说明了乳糖操纵子和组氨酸操纵子的调控机理。*lac* 操纵子系统为诱导型，需要一个不能激活阻遏物的分子(诱导物)诱导。而 *his* 操纵子系统中平时有一个无活性阻遏物存在，当此阻遏物同一个激活它的辅阻遏物结合形成复合体后即可阻止 *his* 操纵子的转录。*lac* 操纵子的诱导物是乳糖(*lac*)，它是操纵子中结构基因 *z* 所编码的 β-半乳糖苷酶的底物。而 *his* 操纵子的辅阻遏物是组氨酸合成代谢的"末端产物"，即组氨酸。(引自 Watson, 1970。)

His-tRNA 的结构发生改变，二者之间存在着必然的联系。另一种突变也已分离获得，表现为使激活组氨酸的合成酶（系指激活组氨酸形成His-tRNA 的合成酶——译注）数量降低，这一突变实际上同改变 tRNA 结构的突变在表型上是相同的。

如同 *lac* 操纵子一样，*his* 操纵子也属于负反馈调控范畴。这二种场合都是依赖于阻遏物同操纵基因结合来抑制转录（图189）。与此相反，另有少数调节机制则属于正（反馈）调控，其中之一为阿拉伯糖系统。在这个系统中不存在阻遏物，只有激活物或起始物。激活物由阿拉伯糖系统的调节基因编码，它同启动结构基因的转录有关，因而对于合成阿拉伯糖代谢系统有关的酶是不可缺少的因素之一（图 190）。

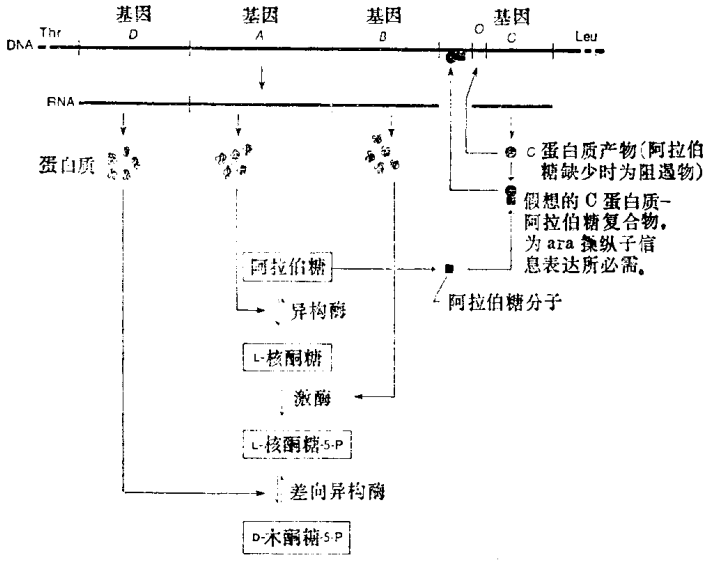


图 190　大肠杆菌中的 *ara* 操纵子

阿拉伯糖（*ara*）操纵子属于正（反馈）控制系统，即由一个起始物诱导 *ara* 操纵子的转录。当缺少阿拉伯糖时，异构酶、激酶和差向异构酶数量很少。若存在阿拉伯糖时，这三种酶都相应增加。这三种酶的增加均受第四个基因 *C*所控制。*ara*操纵子不同于 *lac* 操纵子和 *his*操纵子，其*C*基因的产物是一个激活物而不是一个阻遏物。在缺少这一激活物的场合，酶合成受阻。（引自 Watson, 1970。）

真核生物的调控

在高等生物中，有关蛋白质合成的调控机理迄今尚鲜为人知。细菌中所发现的诱导形式和阻遏形式并未在高等生物中发现，因而不能肯定是否在高等生物中一定存在这样一些机制。在细菌中行之有效的某些遗传学和生物化学的实验，对于高等生物也许并不适用。尽管如此，人们还是知道在高等生物中所存在的许多特殊的调控类型，虽然这类调控并不直接控制蛋白质的合成。本节仅略叙其中一例。

在哺乳动物中，激素作用是极其错综复杂的一种反馈网路体制。Earl Sutherland 新近的工作对理解激素作用的化学本质具有深刻的意义。作者证明，在激素的调节作用中有一种小分子化合物〔环状 AMP(cAMP)〕参与其事（图 191)。Sutherland 最初在描述 cAMP 的作用时曾经提出，cAMP 是旨在接受胰高血糖素和肾上腺素激素的刺激而作为一系列糖元分解反应的中间媒介。糖元的分解反应使糖元分解成葡萄糖，这一反应共经历三个步骤(图 192)。

肾上腺素和胰高血糖素的功能为刺激血糖水平的增加，这同胰岛素的作用正好相反。一个动物随时随地都会有剧烈的肌肉运动发生，因而需要增加血液中葡萄糖的含量，而这种要求是通过胰

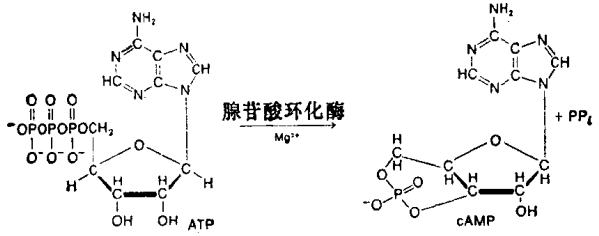


图 191　环化 AMP（cAMP）

腺苷酸环化酶(cAMP 合成酶)可将ATP转变为cAMP和焦磷酸(PP_i)，这个反应需要有 Mg^{2+}参加，并生成二个无机磷酸。酶催化在核糖 5′碳原子上的磷酸通过酯键同 3′碳原子相连，这是一个不可逆反应。另外有一组环化核苷酸磷酸二酯酶可将 cAMP水解成 5′-AMP。此酶一般存在于细胞质中，因此细胞内的cAMP半衰期甚短。

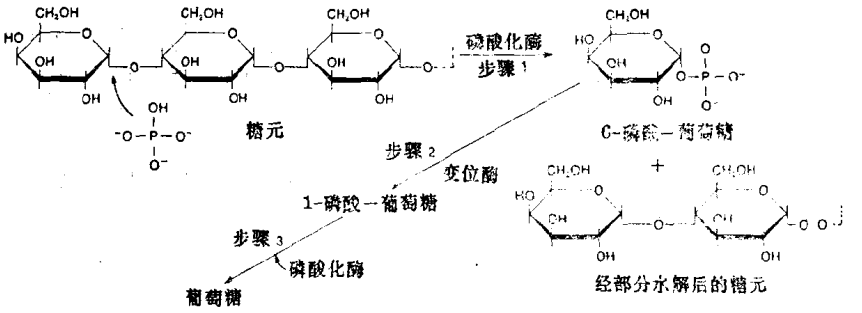


图 192　糖元分解

肾上腺素和胰高血糖素可加速糖元分解以增加血糖水平。糖元分解成葡萄糖包括三个反应步骤。第一个反应步骤(1) 可限制整个的反应速率。胰高血糖素和肾上腺素则通过增加细胞内cAMP的含量来激发磷酸化酶的活性。

高血糖素的分泌来满足的(或者当存在诸如疼痛、愤怒或恐怖等等神经刺激时，由肾上腺素行使这一功能)。这二种激素在化学组成上毫无共同之处(图 193)，但二者都能通过加快糖元分解速度，起到大大提高血糖水平的功效。此外，这二种激素还能引起细胞中 cAMP 浓度的迅猛增加。

胰高血糖素是一种大小约 3,500 道尔顿的小分子蛋白质，象胰岛素一样它也是由胰腺中的胰岛(Langerhans)细胞所分泌。而肾上腺素 (3,4-二羟-α-〔(甲氨)甲基〕-苯乙醇) 则通过交感神经

图 193　胰高血糖素和肾上腺素

肾上腺素和胰高血糖素的化学结构大相径庭，但它们对糖元分解却表现出相同的生理效应。肾上腺素还有一个通用名称，即为 adrenalin。

系统中的神经元，由肾上腺(位于肾脏上部)内部的肾上腺髓所分泌，接受神经刺激并担负产生和分泌肾上腺素的是肾上腺髓中的嗜铬细胞。

贮藏在肝脏和骨骼肌中的糖元在胰高血糖素或者肾上腺素大量分泌时会很快耗尽。这二种激素的作用都在于增强糖元分解中的第一个酶，即糖元磷酸化酶的活力(图 192)。值得注意的是，所谓的磷酸化效应只是一种活化现象——不是增加这类蛋白质的合成，而这种活化作用同 cAMP 含量的增加直接有关(图 194)。

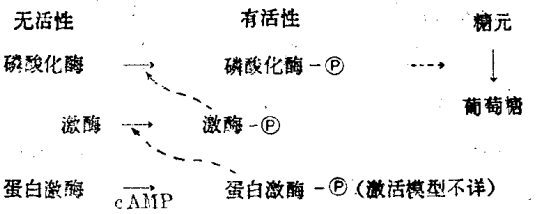


图 194　磷酸化酶激活

图示使糖元磷酸化的磷酸化酶被激活的反应顺序。其中蛋白激酶由 cAMP 激活的机制不明，仅知磷酸化酶的磷酸化对糖元分解是不可缺少的。这里所示为磷酸化酶磷酸化所必需经过的一系列磷酸化反应，整个反应系列起始于蛋白激酶的磷酸化(见最下一行)。

cAMP 含量的升高总是伴随着磷酸化酶活力的增强。为了阐明这二者之间的因果关系，必须弄清在 cAMP 触发下所产生的究竟是怎样的一条反应链。这里，首先一个问题就是从无活性的磷酸化酶转变为有活性的磷酸化酶的磷酸化过程是如何进行的。起初，人们以为激活作用是由cAMP本身将磷酸化酶磷酸化的结果。现在已经知道，磷酸化酶的激活反应实为另一个激酶所催化，而这个激酶自己也需要先磷酸化才具有活性。由于发现了激酶，人们曾设想或许 cAMP 的作用是使激酶磷酸化。可是出乎意料，催化磷酸化反应的却是另一个原先称为激酶激酶的酶（现称蛋白激酶)。蛋白激酶同样也必须预先活化。所以，cAMP 实际上只是一个媒介因子，在它的激发下引起一连串的激活反应。关于这些激活反应，广为流行的解释是：cAMP 通过同蛋白激酶结合并引起后者构象改变而被激活，蛋白激酶的这一变化则又招致由 Jack 提出

的级链式反应成功起动。整个过程包括:cAMP 激活蛋白激酶,蛋白激酶催化激酶磷酸化反应使激酶活化,然后激活了的激酶再使磷酸化酶磷酸化从而使磷酸化酶活化,最后激活了的磷酸化酶催化糖元转变成葡萄糖的第一步反应(图 194)。

虽然人们对这一解释的合理性抱有很大兴趣,但这些反应还只是回答了有关磷酸化酶如何激活的问题,这只不过是整个课题的一个方面。至于另一个方面,即 cAMP 含量的增加是如何诱导的问题,则要靠寻找其他的反应来给予解答。而这些反应实际上并不太复杂,简而言之,即是由腺苷酸环化酶催化 ATP 生成 cAMP,而另一个酶(磷酸二酯酶)则将cAMP水解为5′-AMP。某些能引起 cAMP 含量增加的激素,其作用就在于促进腺苷酸环化酶的活力,或者降低磷酸二酯酶的活力。这二条途径均能使 cAMP 的浓度发生改变。当 cAMP 大量生成时,糖元分解作用即被激活;与之相反,细胞中 cAMP 含量转低时,糖元分解作用便将减慢或停止,血糖水平亦随之降低。

胰高血糖素和肾上腺素可能都是通过与腺苷酸环化酶结合的方式使酶的构象发生改变,结果引起腺苷酸环化酶活力大大增强。这二种激素在化学结构上彼此各异,但生理功能却如出一辙。对于这一困惑不解的问题,从大量资料来看可提出解释如下:在体内存在有二种可将 ATP 转变为 cAMP 的腺苷酸环化酶,一个由胰高血糖素激活,另一个由肾上腺素激活,二种激素的效应是累加性的。

根据多年来的工作,已经提出了关于激素作用的一种模型。这个模型要求在信号的输入部位有一个细胞受体,并有一个特异的细胞反应作为细胞受体接受信号的标志。细胞上的受体部位就是识别激素的位置,刺激(或信号)同受体部位结合后,引起细胞作出反应。cAMP 的作用机理正适合这种模型。例如,胰高血糖素和肾上腺素首先同受体上的腺苷酸环化酶结合,马上细胞作出反应,cAMP 含量即随之增加。从信息流的观点来看,激素应属于第一信使,而 cAMP 则属于第二信使(图 195)。

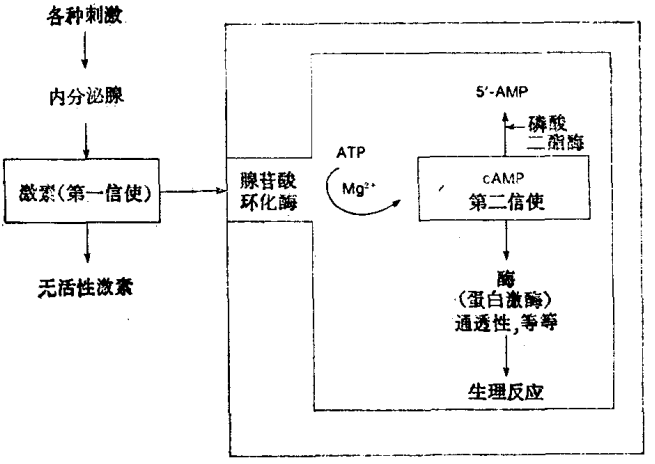


图 195 激素-cAMP 调控机理略解

神经系统发出一个信号(刺激)，诱导内分泌腺分泌激素。这是对刺激作出的第一个生理反应，又叫第一信使。第一信使激活腺苷酸环化酶，诱发第二信使(cAMP)生成。第二信使激起一连串的磷酸化反应，最终使糖元分解成葡萄糖并(或)引起其他生理反应。(引自 G. A. Robison, R. W. Sutherland 与 E. W. Sutherland, Cyclic AMP, Academic Press, New York, 1971。)

目前已经知道，有十多种激素可通过激发腺苷酸环化酶而将它们携带的信息传递给细胞（表 28)。至于每一种激素的独特作用，除胰高血糖素和肾上腺素外其他都不甚清楚。人们往往把激素本身的作用简单地理解为是由靶组织（即接受刺激的组织——译注)的特异性规定的，可是这种认识并不能解释为什么儿茶酚胺可激发多种组织中的腺苷酸环化酶。另一方面，为什么大鼠的脂肪组织可对多种激素作出反应。总而言之，在能够分离得到细胞中各个特定的受体部位之前，人们无法确定细胞中是否存在有专一性各不相同的腺苷酸环化酶；或者腺苷酸环化酶虽然只有一种，但赖以寄付的细胞受体却有多种多样。在哺乳动物的调节机制方面所揭示的胰高血糖素、肾上腺素和腺苷酸环化酶之间的相互关系，已经为人们提供了有关激素作用的新见解。遗憾的是，我们对于激素--受体-反应三方面的相互关系及其特异性的协调作用等机

表 28 某些以 cAMP 作为中间媒介的激素及其效应

靶 组 织	激素(一种或多种激素)	激 素 效 应
脂肪组织	儿茶酚胺, ACTH, 胰高血糖素, TSH, LH, 胰岛素, 前列腺素	加速脂肪分解
肾上腺皮质	ACTH	促进甾类化合物生成(分泌糖皮质激素)
黄 体	LH	促进甾类化合物生成(分泌黄体酮)
卵 巢	LH	促进甾类化合物生成(分泌雌激素和黄体酮)
睾 丸	ICSH(LH)	促进甾类化合物生成(分泌雄激素)
子 宫	儿茶酚胺	促使肌肉松弛
心 脏	儿茶酚胺, 胰高血糖素	加强心肌收缩力
肝	儿茶酚胺, 胰高血糖素, 胰岛素	加速糖元异生作用, 释放 K^+, 促使尿素形成
蛙皮肤	α-MSH, α-类肾上腺素(功)能药物, melatonin	促使黑色素弥散
蟾蜍膀胱	抗利尿激素	促使水及无机离子转运
肾	抗利尿激素, PGE_1	促使水及无机离子转运
肾, 骨	甲状旁腺(激)素	增加血液中钙的含量
腮 腺	肾上腺素	促使分泌淀粉酶
胰 脏	胰高血糖素, α-类肾上腺素(功)能药物	促使胰岛素分泌
某些组织	儿茶酚胺, 胰高血糖素, ACTH, LH, 抗利尿素等等	加速糖元分解

ACTH = 促肾上腺皮质激素; TSH = 甲状腺刺激激素; LH = 促黄体激素; PGE_1 = 前列腺系; ICSH = 促间质细胞激素; α-MSH = 促黑激素。

理大都不甚了然。

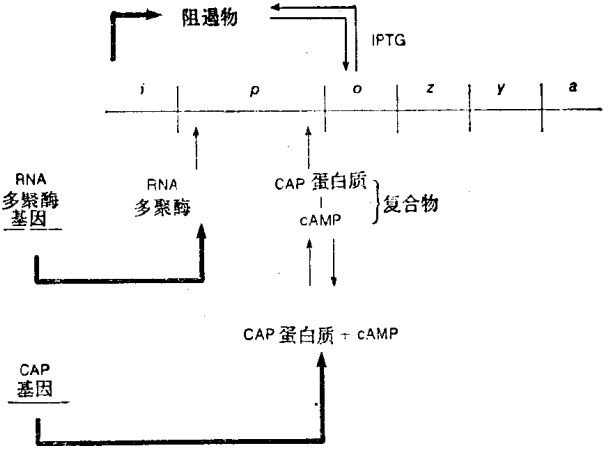
在大肠杆菌中, CTP 调节 ATC 酶的作用机制同哺乳动物中激素引起腺苷酸环化酶的活化有若干相似之处, 二者都能引起某个酶的构象改变。因此人们进而设想, 腺苷酸环化酶很有可能位于细胞的质膜上。假若腺苷酸环化酶的调节亚单位所在位置是朝

向细胞外的体液并且面对激素流经的过道，这样激素一旦附着到调节亚单位上，便使与调节亚单位相对而面向细胞内部的催化亚单位激活，从而引起细胞内含物的反应。当然，这个模型的真实性倒底如何，还有待于设计新的实验来给以验证。

cAMP 的发现曾经引起生化学家、生理学家和遗传学家的广泛注意，它在各种细胞和组织中都有存在。分子遗传学家不久前获证，cAMP 对于大肠杆菌中许多诱导酶的合成也是必不可少的因素。但是不应把 cAMP 与诱导物等量齐观，它与诱导物的不同之处在于，诱导物一般只对特定的操纵子起作用，而cAMP则能影响多种蛋白质的合成。例如，在 *lac* 操纵子中，cAMP 作用于启动位置，以某种方式协助转录的起始。曾经在大肠杆菌中选择了一种不能合成cAMP 的突变型，它们不能在乳糖和半乳糖的培养基上生长，也不能利用其他糖类作为碳源。当在培养基中加入cAMP之后，该突变型的生长能力即便恢复，并能够利用上述任何一种糖源，亦能在含有上述所有糖源的培养基上生长。

试验证明，cAMP 在某些启动子突变型中无活性表现，但在调节基因和操纵基因突变型中则显示活性而不受影响。因此看来，cAMP 的作用机理似乎是通过增加启动转录的频率来激发 β-半乳糖苷酶 mRNA 的合成。图 196 所示即为 cAMP 作为酶合成激活物的一个作用模型。据推测，cAMP 并非直接作用于启动基因的 DNA，而是先同某个受体蛋白形成复合物作为中间媒介，然后受体蛋白-cAMP 复合物再同启动基因 DNA 结合导致转录“开放”。

上面谈的是关于 cAMP 通过什么机制激发蛋白质合成的问题。现在要问，cAMP 在细胞内的浓度是怎样进行调节的？人们发现的第一种调节方式是直接控制法，即通过抑制或激活 cAMP 磷酸二酯酶的活力来行使调节作用（图 195）。第二种方式（只在细菌中被发现）是将细胞内的 cAMP 转送到细胞外的培养基中以降低体内 cAMP 的浓度。第三种方式则是借助于控制腺苷酸环化酶的活力用以调节 cAMP 的生成。在大肠杆菌中，葡萄糖也能



不同处理	*lac* mRNA 的相对合成速率	*lac* mRNA 的相对浓度	β-半乳糖苷酶的合成速率
空白	7	4	1
IPTG	100	100	100
IPTG+葡萄糖	12	10	15
IPTG+葡萄糖+cAMP	119	82	130

图 196 cAMP 对 *lac* 操纵子的影响

cAMP 影响 *lac* 操纵子机理如下：首先 cAMP 与 CAP 蛋白质(cAMP 受体蛋白质)结合，然后复合物同启动子区域(*p*)结合。操纵子的*p* 区是识别 RNA 多聚酶的部位，而启动 *lac*-mRNA合成的是 cAMP-CAP 蛋白质复合物。下表所示分别为诱导物(即IPTG，其化学结构与乳糖类似，可同阻遏物结合但不能作为能源利用)，诱导物＋葡萄糖，诱导物＋葡萄糖＋cAMP 三种处理的相对效应以及 *lac* 操纵子的活力表现。IPTG 只激发 *lac* 操纵子的活性，而葡萄糖又能使这种效应消失，这是因为当乳糖诱导物不存在时，葡萄糖本身可作为能源利用。葡萄糖对IPTG产生的抑制效应可由cAMP予以消除。(引自I. Pastan 与R. Perlman, Science, 169, 339, 1970。)

抑制腺苷酸环化酶的活力，使 cAMP 的含量降低。有许多正常的细胞代谢产物都表现有抑制腺苷酸环化酶活力的作用，另有一些代谢产物则有促进该酶活力的作用。

目前，有关这方面的研究范围已愈来愈广。十年前，人们又发现一个环化核苷酸，即 cGMP。它似乎以 cAMP 相反的作用方

式影响细胞和器官内的代谢进程，因而颇为引人注目。例如，当cGMP 含量迅速增加的同时，由乙酰胆碱诱导的心脏收缩力即相应降低。cGMP 对于心脏的这种作用效应与肾上腺素刺激心肌中cAMP 含量增加所产生的效应恰好相反。cGMP 与乙酰胆碱之间存在的此种对应关系在其他组织中也能观察得到。另外，还发现有一些药剂既可以激发细胞内 cGMP 水平的升高，又可同时促进细胞内 cAMP 含量剧增。到目前为止，可以有把握地说，相应于cAMP 的每一个正向作用的发生都可能存在一个准确的由 cGMP 激发的反向作用或拮抗作用相伴随。换句话说，cAMP 的最终控制效应并不完全依赖于它在细胞中的绝对浓度，它还取决于cGMP 浓度的相互制约，即依赖于 cAMP 与 cGMP 之间的浓度比。

具有活性功能的鸟苷酸环化酶已经在许多哺乳动物中找到。采用各种动物和细菌所做的试验均已证明，cGMP 同样渊源于GTP。鸟苷酸环化酶不同于腺苷酸环化酶，在离心提取时它并不存在于颗粒沉淀(膜碎片)组分中，而是存在于细胞提取物的可溶性组分中，因而不能判定该酶在体内的确切位置。迄今为止，亦未发现鸟苷酸环化酶可被其他具有生物活性的物质所激活，但是此酶的许多抑制剂却已早为人知。在这些抑制剂中，唯有 ATP 浓度变化对鸟苷酸环化酶的抑制效应同体内发现的相同。由此看来，这二种小分子(cGMP 和 ATP)在激素调节的网路中彼此牵制，相辅相成，从而使体内代谢平衡达到一个适宜的水平。

在结束代谢调节这一课题的讨论之前，最后还要强调一点。即在上述所有事例中，我们对信息流的描述或只停留在蛋白质已经合成阶段，或在说明蛋白质催化某一反应之后即告终止。至于从蛋白质合成之后到表型显现这中间究竟经历了什么变化过程，仍然并不清楚。在真核生物中尤其如此。这个问题看来更接近于发育生物学所研究的范畴，要想进行实验分析相当困难。主要原因在于，表型是基因型与环境相互作用的结果，它表现为一种综合的效应。对于这样一种综合效应，想要从中分辨出每个参与表型发

育的基因所发挥的各自效应，决非是件轻而易举的事情，特别是那些有关复杂性状的表型，诸如身高、智力或行为等等。在真核生物中，人们目前连蛋白质合成的调控机理都不清楚，初步了解的有关转录和翻译的问题也只局限于少数蛋白质。考虑到真核生物中受调控的蛋白质何止千万，各调控系统间的相互作用又多么复杂，这一切都使得上述课题的解决更是难上加难。就以一个简单的真核类粘菌 *Dictyostelium* 为例，Harvey Lodish 就曾证明该粘菌在不同的发育阶段有着性质全然不同的蛋白质产生，亦即暗示个体发育也是受调控机理指导的(图 197)。

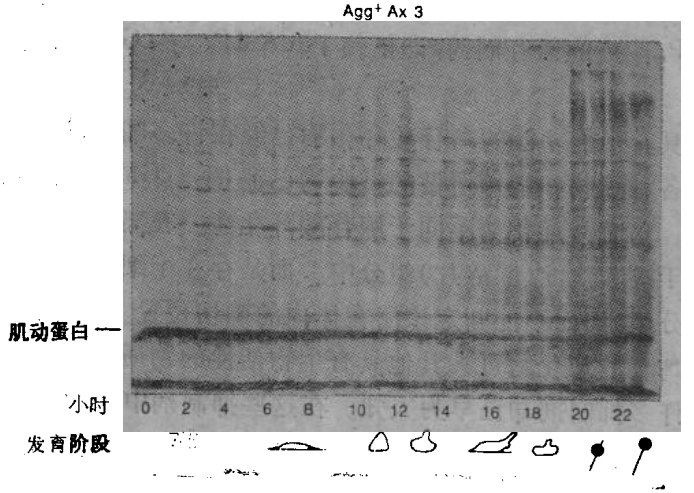


图 197 *Dictyostelium* 生命周史中蛋白质种类变化图

Dictyostelium 是被广泛用来进行发育研究的一种粘菌。它的发育进程经历一系列显著变化：开始时表现为由众多的类似阿米巴单细胞组成的群体，接着相互聚集形成一个棒形的假变形体，最后则发育为一个着生孢子的结构，新的孢子即由此产生。将 *Dictyostelium* 生命周史中不同发育阶段与各个阶段当时产生的蛋白质类群作对比分析，绘制本图。这项观察的含义主要是说明发育进程同样是服从于开-关调节机制指导的，但其详细内容不明。图中的蛋白质不同区带为聚丙烯酰胺凝胶电泳图。(由麻省理工学院 H. L. Lodish赠送。)

即使每一种蛋白质的功能都一清二楚，而且具有它们在机体生命周期中的出没消长的完备知识，但是又怎么样能够测知，在某一特定发育阶段，不同的蛋白质究竟以何种配置组合才会产生所

观察到的特定表型？事实上，随着生物体体制的复杂程度加深，与之相应的信息流花样也必定益趋浩繁。由此看来，明白无误的事实是，生物学中的基本问题尚远未解决，人们无疑将因之会面临一种长期的挑战。

参 考 文 献

Bernhard, S.: *The Structure and Function of Enzymes*, Benjamin, New York, 1968.

Bier, M.: *Electrophoresis*, 2d ed., Academic, New York, 1968.

Dayhoff, M. O., and R. V. Eck: *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Washington, D.C., 1975.

Determann, H.: *Gel Chromatography*, Springer-Verlag, Berlin, 1968.

Dickerson, R. E., and I. Geis: *The Structure and Action of Proteins*, Harper & Row, New York, 1969.

Edman, P.: A Protein Sequenator, *Eur. J. Biochem.*, **1**:80 (1967).

Gutefreund, H.: *Introduction to the Study of Enzymes*, Wiley, New York, 1965.

Gutte, B., and R. Merrifield: The Total Synthesis of an Enzyme with Ribonuclease A Activity, *J. Am. Chem. Soc.*, **91**:501 (1969).

Kendrew, J. C.: The Three-dimensional Structure of a Protein Molecule, *Sci. Am.*, **205**:96 (December 1961).

Koshland, D. E., Jr., and K. E. Neet: The Catalytic and Regulatory Properties of Enzymes, *Ann. Rev. Biochem.*, **37**:359 (1968).

Phillips, D. C.: The Three-dimensional Structure of an Enzyme Molecule, *Sci. Am.*, **215**:78 (November 1966).

Schachman, H. K.: *Ultracentrifugation in Biochemistry*, Academic, New York, 1959.

Schroeder, W. A.: *The Primary Structure of Proteins*, Harper & Row, New York, 1968.

Speckman, D. H., W. H. Stein, and S. Moore: Automatic Recording Apparatus for Use in the Chromatography of Amino Acids, *Anal. Chem.*, **30**:1190 (1958).

Stryer, L.: Implications of X-Ray Crystallographic Studies of Protein Structure, *Ann. Rev. Biochem.*, **37**:25 (1968).

第六章　进　　化

前 生 命 进 化

在一切现代的原生生物、植物和动物种属中，对于遗传信息的贮存、复制、转录与翻译，方式基本雷同。DNA、RNA 以及蛋白质作为信息大分子遍布于整个生物界，而且这些大分子化合物除了初级结构不同之外，任何物种均别无二致。人们完全可以根据这样一种假设——即目前以信息大分子为中心的遗传机制，从历史的观点看不大可能一开始便以现存的形式见诸于世——来提出问题，这就是：各种遗传体制究竟是怎样地从比较简单的形式发展起来的？事实上，这个问题已经概括了分子遗传学所要阐明的学科本身三维性的第三个方面，亦即时间矢量上的历史演化（图198）。

关于以进化这一命题写成的专著已有不下数百种之多，其中多数作品常常以褒桔贬苹的态度，也就是说，以表明圣经可信，企图否定进化史的态度，从而坚持反进化论的宗教式的立场；有些书籍虽然带有朴素进化论的进步的反宗教的色彩，但内容平平也只不过是把桔苹之间已经被颠倒的价值重新颠倒过来而已；第三类著作则真正地把全神贯注于进化论这一主题，详尽地阐释了现代物种的生物学历史。总之，在有关揭示现代物种种属历史的证据方面，人们已掌握有多种多样的材料，因而这类著作涉及的范围极为广泛，这里包括有对古代物种及现代物种进行比较研究的比较解剖学、比较生理学以及生物化学、遗传学等等许多学科。这些内容对于本书来说涉及面实在过大，本章只拟就下述几个问题加以说明：

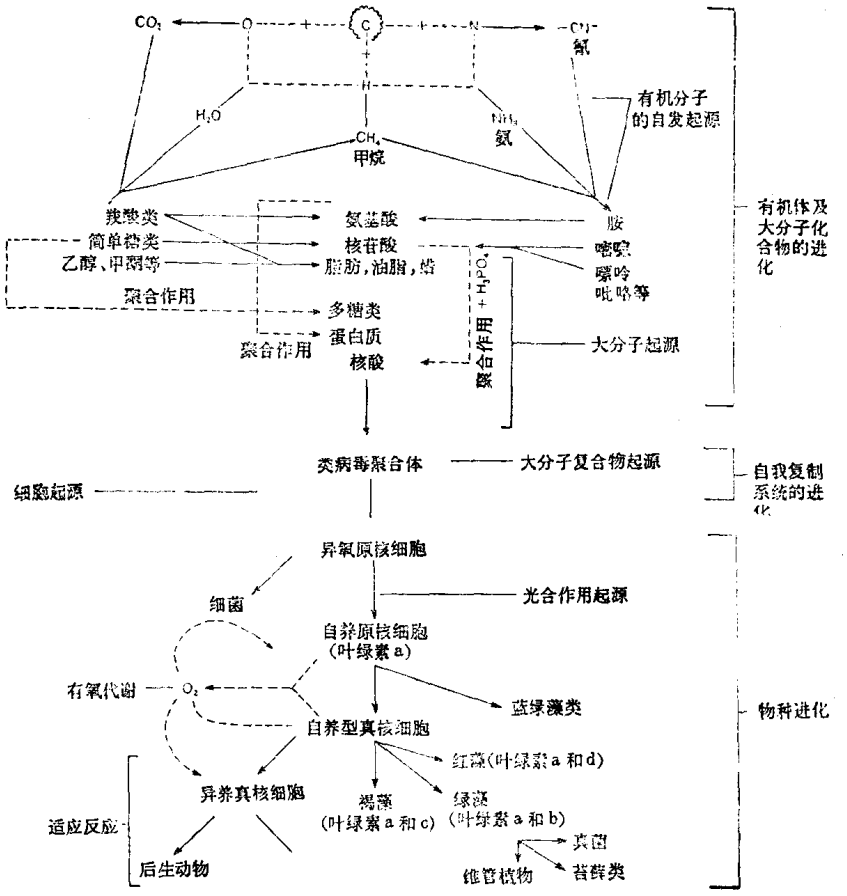


图 198　进化：遗传信息流的时间矢量

从上到下可以人为地把进化划分为三个阶段(见右边括号)。第一阶段包括来自无机物的有机分子和大分子化合物的进化，这就是前生命进化。目前在实验室条件下完全可以做到人工再现上面提出的这一进化道路，但是要想进化历史来一次重演，以便证实或否定前生命进化的假说，则是行不通的。到第二个阶段时，自我复制机制以及各种遗传器等均已演化形成，并日趋完善。自然选择开始对"有机体"直接发生作用。第三阶段是"物种迅速形成"时期，有关研究进化这一时期的文献(包括进化事例与进化过程)甚为浩瀚。对以突变与自然选择作为主要内容的物种进化机制，通过对这一进化阶段的研究亦会有所了解。(引自N. M. Jessop, 1970。)

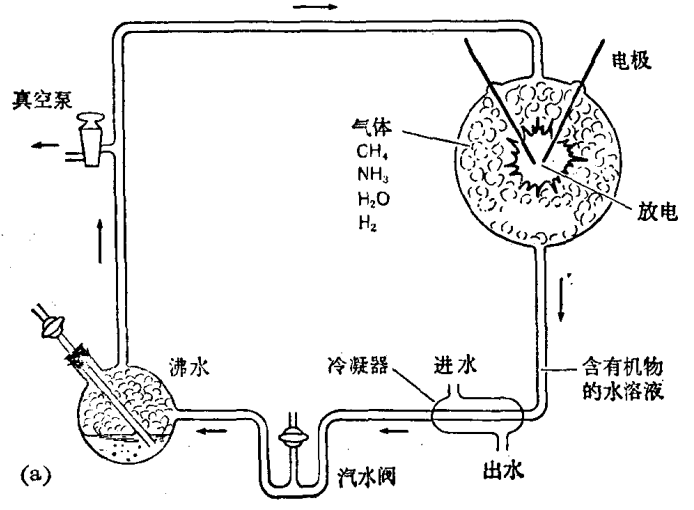
1）有关包括大分子化合物以及带有类似蛋白质膜的微球体在内的“有机”分子前生命进化的某些证据(图 198,上部分)；

2）试图对遗传器进化作推测性探讨(图 198,中部分)；

3）旨在通过对蛋白质及 DNA 初级结构与三级结构的比较研究从分子水平上来阐明物种进化。

在上述诸点中，讨论的重点是进化问题。图 198 中所列的模型只是旨在提出有关各级进化过程的一个概貌，本意并不要求在科学性上一定正确无误,在这张图的各级步骤上,有关的实验证据其实很少或者近乎乌有。在进化过程中有许许多多的生物种属形成和发展了，但是承认这一事实并不等于说人们已经对其中所涉及的全部机制或是所有种属进化的步骤都已有所认识。譬如，我们对遗传器进化知之极少便是一例，这部分是因为历史残留的证据零星不全，此外与人们在遗传器进化过程方面未能提出一种较为明确的概念作为探讨这一问题的思想指导亦不无关系。

如果我们仔细考察一下现代各种遗传器的结构与其作用机理，然后倒转历史追溯它的起源，就会发现无法回避这一问题，即在一个有功能的遗传器发生以前，必然发生有若干看来不大可



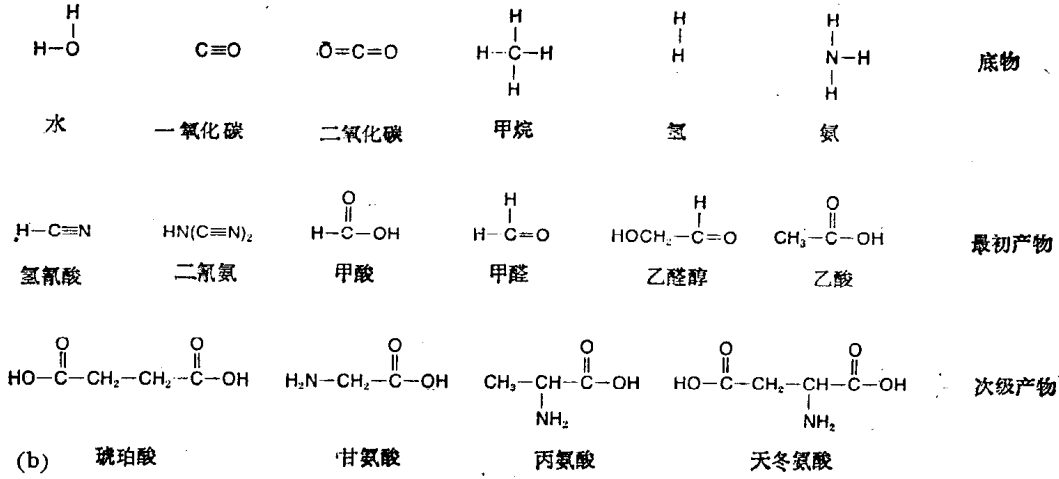


图 199　有机分子的非酶促起源

图b上行所列为导入装置(图a)的无机分子底物的化学结构式。图b中所列为经短暂放电反应后获得的最初产物，即为无机分子和有功能的有机分子之间的过渡产物。如放电时间较长，则产生图 b 下行所列的所谓次级产物的各种分子。这里请读者注意最初产物与次级产物在化学结构上所表现的相似性。(引自 M. Calvin, 1969。)

能发生的事件巧遇和物质组合。从现代遗传体制推想，第一个完整的遗传器大概是处在一个能长期自我保存的微环境(初始细胞)内部，最低限度它必须包括有两个核酸模板(初始基因)和两个相应的酶（一个用于代谢，一个用于复制——译注)。只要掌握了遗传器到底如何发生的知识，即使极其有限，人们亦会因之获得某种进行推论的活动余地，以便谋求建立能据以使进化过程系统化理性化的模型。

根据地质资料和化石的证据，一般公认地球上的生命是在大约三十亿年前开始出现的，在开头二十亿年的进化过程中，尚未能形成广泛变异的细胞类型，达尔文式的物种进化过程的迅猛发展大概是在十亿年前。据估计在这期间产生并繁衍的动植物种属约在一亿种以上，其中有一百多万种一直存活至今。这两个时间点可能是生物进化历史的"里程碑"。在第一个转折点前，由于化学反应及物理变化推动，导致逐渐形成某种"有机物"分子。生命的出现又进一步带来了新的和更为快速的化学反应。生活在过去二十亿年之间的有机体可能由于能量的限制，不如现代物种这样五花八门。这一时期有机体均为厌气性的代谢，氧分子还派不上用处。最后，在有氧代谢所必需的酶和机体结构发展的基础上，加上地球上氧分子广泛形成，物种的多样性已随之急剧增加，并最终奠定了地球上现今存在的纷繁复杂的生命形式。

由此可见，在生命进化史上存在有两个重要的里程碑：一个大约在三十亿年前，为生命起源的时刻，另一个大约在十亿年前，为物种加速繁衍的时刻。这里不妨先探究一下生命起源有可能据以起始的环境因素。苏联生物学家 A. I. Oparin 曾企图重建他所认为的为导致生命起源必备的某些物理和化学过程。Oparin 的理论过于复杂，不便在此详加引述，值得指出的是，根据这一理论所设计的大量实验已经提供了为数甚多的富有意义的资料。这一切对于 Oparin 的理论都是有益的支持。一般有关生命起源的理论都认为：远在有生命的细胞出现之前，各种类型的作为形成蛋白质、核酸、碳水化合物所必需的"有机分子"便已充斥着整个海洋

与湖泊。总之，当时的所谓海水，实际上犹如是充满氨基酸、嘌呤、嘧啶、葡萄糖等等化合物分子的有机浆液，正好为第一个形成的"细胞"准备好现成的食料。顺着这条线索推理，如此这般的第一个"细胞"应该具有完全异养型的特征。

第一个尝试用实验检验 Oparin 的一部分理论而获成功的是 Harold Uvey 和 Stanley Miller 两人(1952)。方法是按照Oparin 理论暗示的成分组成，建造人工大气环境，再在其中安装一枚电极使之与气体分子直接接触(图 199)。实验以甲烷、氨、水及氢气为起始底物，放电反应后，形成的产物包括有多种氨基酸、嘌呤、嘧啶以及一些简单的糖类分子。事实上，上述各类聚合性产物都是在同一人工环境系统中从小分子化合物元件一炉派生的。

根据实验观察结果，可以得出两点明确无误的结论：(1) 小分子有机物能够在活细胞外部自然形成；(2)通常在活细胞中作为大分子化合物结构元件的那些化合物分子，当在给予能源和无机化合物前体的情况下，热力学条件亦有利于它们脱胎而出。换句话说，氨基酸本身并不是在其合成中必须要接受活细胞遗传信息指导的非自主性分子，恰恰相反，只要有前体存在并提供充足的为形成共价键所必需的能量，在体外环境中产生氨基酸亦无多大困难。

表 29　在星际空间发现的分子

分　子	名　称	分　子	名　称
H	氢	CH_3OH	甲　醇
OH	羟基	HCN	氰 氢 酸
NH_3	氨	CHCCN	氰乙烯炔
H_2O	水	HCOOH	甲　酸
H_2CO	甲醛	CH_3CCH	丙　炔
CO	一氧化碳	HNCO	异 氰 酸
CN	氰		

引自 S. Fox, 1972。

表 30 木星大气成分

气　　体	%
CH_4	1
NH_3	0.05
H_2	60
He	36
Ne	5

引自 M. Calvin, 1969。

表 31 由 $^{14}CH_4$, NH_3, H_2 和 H_2O 通过辐射产生的化合物分类

组　　分	产　　量
酸类化合物	45.5
用离子交换树脂 Dowex 1 及 Dowex 50 离析的碱类化合物	17.44
非离子成分	21.3

	在混合产物中的含量%		
	总　量	在碱性组分中的含量	在酸性组分中的含量
氰氢酸	0.45		
腺嘌呤	0.34	0.203	
5-氨基咪唑甲酰胺	0.018	0.105	
乳　酸	0.99		0.21
两种酸性占优势的未知化合物（不含乳酸）			29.0和17.8
甘氨酸	0.03	0.2	
α-丙氨酸	0.18	1.07	
天冬氨酸	0.03	0.2	
两种碱性占优势的未知化合物（茚三酮-阳性，但亦非尿素或胍）		20.9及11.9	

引自 M. Calvin, 1969

表 32 CH_4, NH_3, H_2O, H_2 混合物的放电产物

化合物分子	产量(分子数×10^5)	化合物分子	产量(分子数×10^5)
甘氨酸	63		
羟基乙酸	56		
肌氨酸	5	天冬氨酸	0.4
丙氨酸	34	谷氨酸	0.6
乳　酸	31	亚氨基二乙酸	5.5
N-甲基丙氨酸	1	亚氨基丙乙酸	1.5
α-氨基-n-丁酸	5	甲　酸	233
α-氨基异丁酸	0.1	乙　酸	15
α-羟基丁酸	5	丙　酸	13
β-丙氨酸	15	尿　素	2.0
琥珀酸	4	N-甲基尿素	1.5

达尔文虽然没有考虑过有关分子进化的细节，但他确曾预言，生命的产生和生命形式的发展毋需等大自然将各种化合物分子在功能和结构的全部排列组合都一一排演之后才得以进行，一般认为，在大量的理论上可能的化合物(或元素)组合中，能由无机底物自然形成者只占其中少数，这一情况与 Miller-Urey 的实验结果也大体吻合。

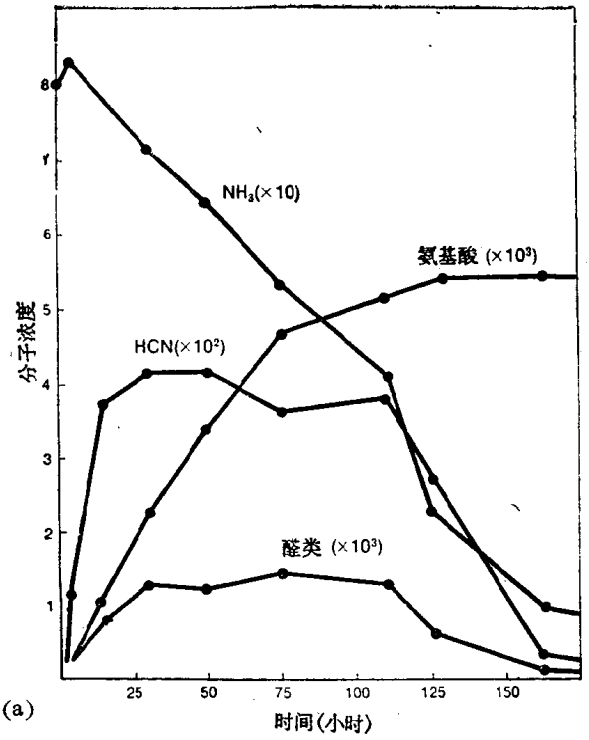
Miller 和 Urey 实验中选用的无机底物分子并非随机组合的样品，而是为模拟重建 Oparin 理论中提出的前生命大气环境而严格选择的。Oparin 的主要理论最早发表于 1936 年。1952 年，Miller 和 Urey 报道了自己的实验。从 1968 年伊始，人们发现，在星际空间(表 29,30)同样存在有与 Oparin 暗示的并由 Miller-Urey 着手实验验证的完全类似的无机化合物分子。除此以外，还曾宣布在太空陨石以及月球尘埃中甚至含有氨基酸成分的证据，不过由于其中氨基酸的含量过低，因此认为不能排除污染的因素，从而使这类报告的真正价值受到质疑。

为了再现前生命进化时期的地球面貌，曾经模拟了各种各样的人造环境，为数甚多的有机化合物分子均在这类人工条件下从无机分子合成成功，例如，丙炔氰转化为嘧啶，HCN 经五聚化作用产生腺嘌呤(表 31，32，图 200a、b)等等，在其他各种不同条件下则相应可以产生尿嘧啶、胸腺嘧啶和鸟嘌呤，甚至多肽也能在同样的人工环境中形成(图 200c)。根据这类实验，提出如下推理不见得便荒诞无稽：如果给以水分、氨气、甲烷、氢气以及能源与多种多样的无机、有机化合物分子，由之自然产生大分子化合物乃至最终出现生命系统将是合乎逻辑的结果。

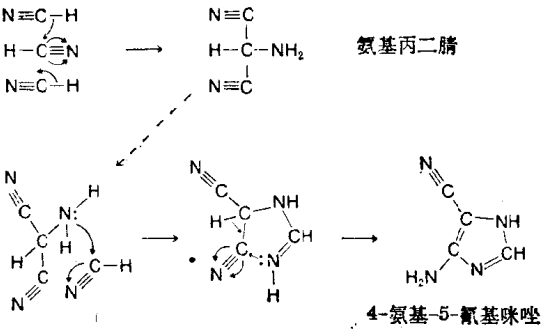
将氨基酸混合物加热，当温度达到 140～180℃时即可获得一些长度约在 50～90 个氨基酸残基之间的肽链片断 (Sidney Fox 曾在这方面做了大量工作，他把这类多肽称做类蛋白)，由于类蛋白的一级排列顺序没有专一性，结果人工获得的类蛋白分子也千变万化。这些类蛋白分子之所以引人入胜，是因为某些类型的类蛋白分子居然具有催化(诸如水解、脱羧，氨基化、脱氨及氧化还原等)各种反应的能力，有的类蛋白能和亚铁血红素分子的一半结合产生亚铁血红素类蛋白，另外一些类蛋白则表现有过氧化酶活性(表 33)。类蛋白催化反应的活性水平同现今由遗传编码的酶分子相比虽然望尘莫及，但是一旦有催化作用发生，不管强度如何，它必定会促使提高某些化学反应产生的概率。

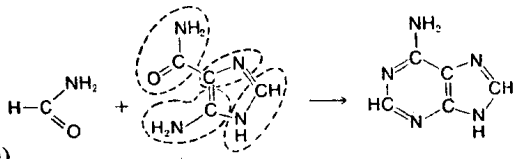
那些对 Oparin 的论点作过检验并试图重建前生命地球环境的作者都一致同意，地球当时的表面气温很可能在 140～180℃之间，事实上，在美国，地表面温度达到 100℃甚至更高者现今还有一千多处。如果考虑到在高温条件下只需要几个小时便能完成聚合反应，因此，可以想象，即使温度在 140℃以下发生聚合的可能性同样也会很大。

在有关前生命进化的理论中，多数作者都强调了建立一个封闭性的类细胞环境的重要性。有了这样一种封闭系统，才能使一定种类的化学反应过程与"细胞"外部发生的化学反应过程相互隔绝。如果新发生的特定反应能够做到跟细胞外环境中的各种偶发


NH₃(×10)
氨基酸 (×10³)
HCN(×10²)
醛类 (×10³)
分子浓度
时间(小时)
25
50
75
100
125
150

(a)


氨基丙二腈
4-氨基-5-氰基咪唑


4-甲酰胺-5-氨基咪唑
腺嘌呤

(b)

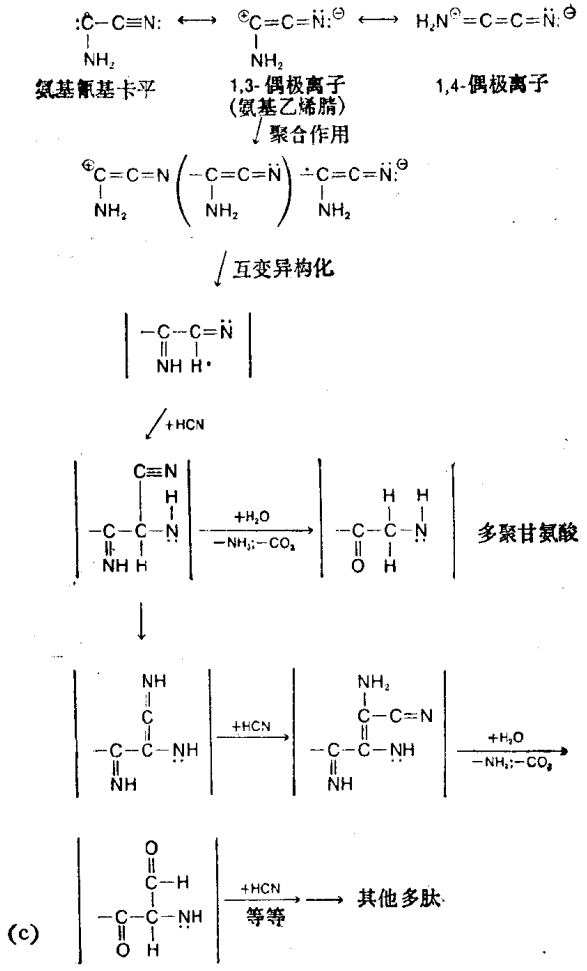


图 200 Miller-Urey 装置中化学反应的种类和速度

(a) 图中曲线表示氨、氰氢酸、醛、氨基酸的产生浓度与反应时间的关系。反应开始时，氨浓度最高(初始产物)，随着氨气消耗，另外三种化合物分子即逐步增加，最终，醛、氰氢酸的浓度达到了顶峰(平衡)，接着，氨基酸浓度不断升高，与之相应，醛和氰氢酸的含量又趋向减少。(引自M. Calvin, 1969。)

(b) 腺嘌呤形成途径中的大部分步骤已经在实验室条件下得到证实，但是有关甲酰氨基与氨基咪唑甲酰胺反应产生腺嘌呤这一最后步骤是否也能在实验条件下发生迄今尚无定论。(引自 J. Oro, 1961。)

(c) 这是一个假想性的机理，旨在解释前生命进化中的多肽形成，过程中包含着两点基本假定：第一，肽链形成是在缺乏肽键合成酶及 mRNA 转录样板的情况下进行的；第二，在某些场合，肽键可先于氨基酸形成，也就是说某些残基能在肽键形成之后转换(加工)为氨基酸。当然，上述机理只是在生命前环境中才有重要意义。(引自 A. Akabor, 1955。)

表 33　在加热条件下脱水聚 α-氨基酸(类蛋白)的催化活性

反应类型与底物分子	反　应　特　征	报告年分
水解:		
p-硝基苯乙酸	类蛋白的活性大于等量自由组氨酸的活性	1962
p-硝基苯乙酸	加热时聚合体活力最高	1962
p-硝基苯乙酸	能被有机磷抑制;亦能催化逆反应	1967
p-硝基苯乙酸	一般描述(无特别新的内容)	1967
p-硝基苯乙酸	具有反应位点并能钝化	1967
ATP	通过 Zn 盐起作用	1965
p-硝基苯磷酸	第二位磷酸水解	1968
脱羧作用:		
葡萄糖醛酸	从葡萄糖、CO_2 催化而成	1964
丙酮酸	生成乙酸 + CO_2,米氏动力学	1966
草酰乙酸	反应迅速,需要碱性多聚体分子	1967
氨基化:		
α-酮戊二酸	同时需要 Cu^{++}及类蛋白存在	1967
脱氨基作用:		
谷氨酸	同时需要 Cu^{++}及类蛋白存在	1968
氧化还原作用:		
H_2O_2(触酶反应)	当氯高铁血红素结合进类蛋白时其活性降低	1971
H_2O_2 及H原子供体(氢醌,NADH 及其他化合物分子)(过氧化物酶反应)	在富赖氨酸的类血素蛋白中氯高铁血红素的活性增高达 50 倍	1971

事件互不相通，这样一种微环境条件也就事实上在起作用了。按照 Fox 的看法,类蛋白分子的特性之一就是它们能以球体或微球体的形式构成有效的微环境(图 201)。微球体结构稳定，大小一致,具有渗透性,与细菌中发现的微球体表现相同的染色性质，并看来能够以增添出芽、分裂的方式进行生长繁殖。很明显,犹如一般所说的自动装置的机理一样，它们大概也是类蛋白分子所固有

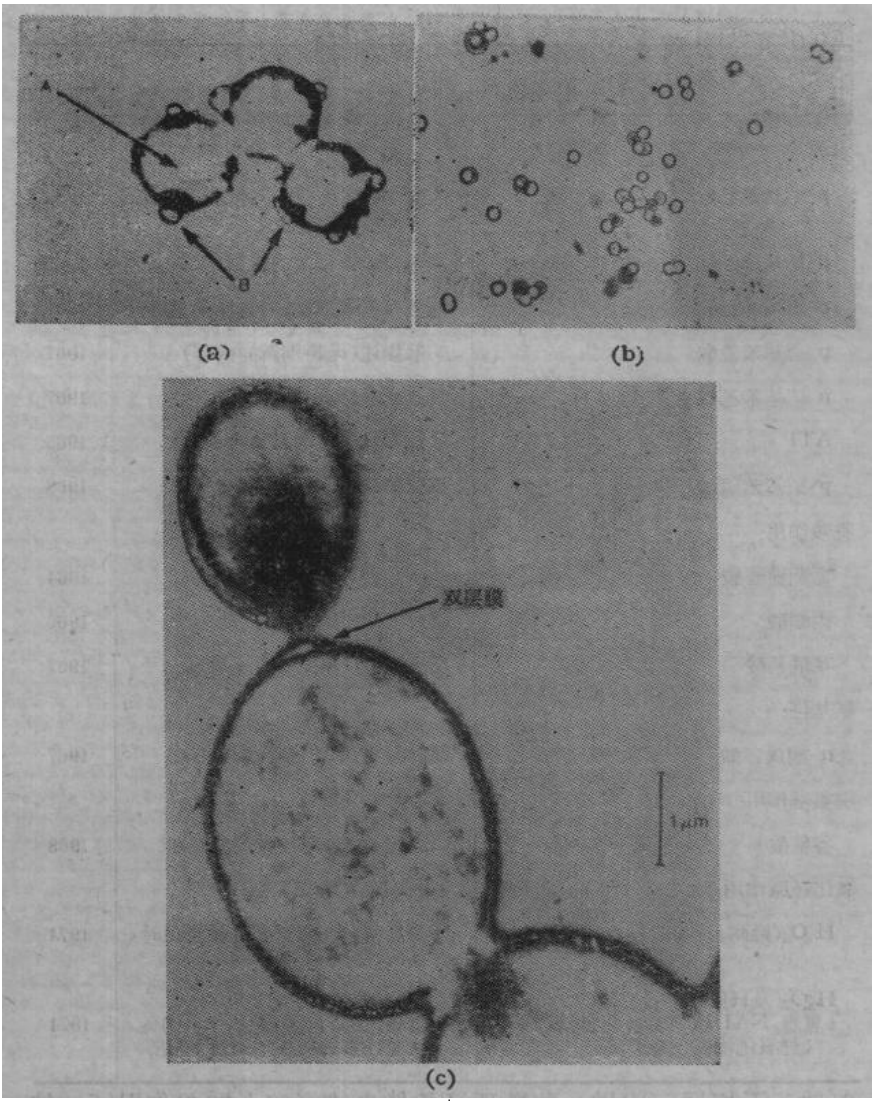


图 201 类蛋白分子自动装配成微球体

(a) 亲本微球体(A)似乎正在出芽(B)。
(b) 各种大小微球体的低倍放大图像。
(c) 双重膜微球体的电镜摄像。据 Fox 记载，微球体具有如下特征：(1)呈稳定状态；(2)大小介于 1～10nm (直径)之间；(3) 易于迅速形成；(4)具有类似于细菌的革兰氏阳性和革兰氏阴性的染色特性；(5) 有半渗透性膜，在清水中表现出渗透性；(6)有催化活性；(7)以扩增和出芽方式生长；(8)具有特定的内含物即核类蛋白质。(由 Sidney Fox 惠赠，选自 S. Fox 与 K. Dose 所著 Molecular Evolution and the Origin of life. Freeman. San Francisco, 1972.)

的自动装配特性的产物，不过对微球体来说，自动装配的对合特异性不甚严格罢了。

我们说，没有任何人真正知道在前生命进化期间实际上到底发生的是什么事情这一点是正确的，可是从另一方面来看，根据Miller，Urey 和 Fox 的实验资料，我们说大分子化合物和微球体均有可能起源于无机化合物分子，这同样也是正确的。目前在此类研究中手头缺少的还是旨在论证遗传器可能来源于前生命进化过程中的大分子化合物和微球体结构这方面令人信服的资料。为了如实地反映我们对进化过程的设想，读者不妨再看一看图198中所展示的全部进化模型，图中，在前生命进化与出现自身复制的功能系统之间有一段空白，这就是上面所说的那一个“缺失环节”。

遗传器的进化

科学家们已经写下了大量著作来解释无机分子进化和物种的迅猛发展之间所存在的这一段空白。纵观整个生命史，有一个问题特别突出(图 202)，人们要问：大约十亿年前引起进化速度迅猛变化的转机到底是什么？有不少研究进化的学者认为之所以发生这种情况主要是由于好气代谢的推动。这一看法确实是引人注目的，如果人们将之理解为物种形成仅仅只是受能不能产生有效的能量发生器所限制的话，但是，如在早期讨论中便提出的，好气代谢需要一套极其复杂的，在结构与功能上能够配合协调的酶系统，在没有遗传器存在的场合，无论如何也难以想象此种复杂酶系统的发展径路，因为这一体制(必定只有)通过突变才能灵敏地适应环境，从而也才能够在自然选择过程中得到保存和发展。

下面讨论的是另一些证据。资料表明，电子传递链中的某些酶在进化过程中没有发生过大的变化，人们发现，几乎在所有的真核类生物中这些酶都执行同一的功能，而且在已经研究的生物种属中其结构也大体雷同，由于酶蛋白的一级结构取决于基因的一级结构，因此可以想象，与这些酶蛋白相对应的基因必定也是惊人地稳定。从现代生物种属中所见到的这类酶的保守性来看，一旦

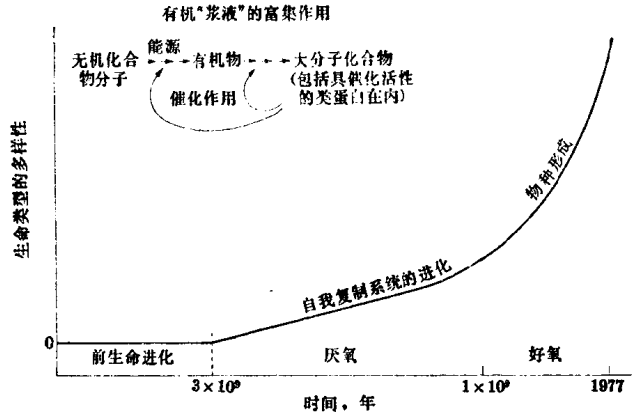


图 202　三大主要进化时期

第一个主要进化时期(前生命时期)标志着"有机浆液"的形成阶段，亦即包括大分子化合物在内的非酶促形成有机化合物分子。第二个时期为生命的起源，是自身复制系统的进化。第三个时期不易与第二个时期明确分开，这是物种迅猛形成的阶段。

好气代谢进化发展之后，有关的酶即保持相对不变。

然而，争论的问题远比上面说的要复杂得多，例如，设若好气代谢是导致物种迅速进化的重要因素，则物种形成的迅猛程度势必会与分子氧的得以利用两相符合，看来，厌氧代谢先于有氧代谢这一点是明白无误的。考虑到分子态氧根本上来自于光合作用，因此即可推论光合作用也必定要比好氧代谢更为久远。不过并非所有的光合作用过程都能够产生出氧气分子，譬如现存的某些细菌中便表现有无氧性的光合作用，它们不象在有氧光合作用中那样以H_2O分子作为电子受体，而是以H_2S、CH_4以及NH_3分子作为电子受体。而所有这些电子受体在当时那种能够产生前生命汁液的还原性大气环境中，都是不难获得的。我们现今还不大可能来科学地描绘有关氧分子产生之前光合作用系统的进化过程，而一般地把氧分子起源归之于光合作用则比较言之可信。

若光合作用先于有氧呼吸的进化过程而存在，便可由此得出如下推论：即高度复杂的基因——酶的相互关系也比好氧呼吸出现的更早。确实，光合作用和有氧呼吸的电子传递系在许多方面

都如出一辙，所以，很有可能有氧呼吸过程中的酶系乃是光合作用中酶系的进一步发展和衍伸。尽管对好氧呼吸的进化还不是一目了然，但可以肯定，在氧气分子可以得来以前，一个有功能的遗传器必定便已经在起作用了，并且随着有氧呼吸机理的进化，必将带来的光合作用的特征性转变以及从厌氧代谢到有氧代谢的发展，很可能这些过程本身已经包含着要求对于电子传递系中的酶系统作相对微细的改变和有效地保存此类变异。

如果有必要排一排各种事件发生的前后顺序，那么，遗传器的进化在时间上的位置肯定是在前生命进化末期与光合作用机制的初始之间，换言之，亦即在前述的第一次转折之后到光合作用系统已经发育完好的机体起源之前，此类遗传器即已在进化过程中出现。

为了探讨遗传器的进化问题，这里需要先就核苷酸和氨基酸之间的初始编码关系以及复制和翻译二个用语的含义作一说明。首先我们要问：原始的遗传器（接合体分子、信使分子和核糖体分子的复合物）的各个组分是独立地完成演化然后在某些点上分别及时地组装成为一个完整遗传器的呢？还是该遗传器的进化以某种功能上整体协调的方式进行，从而使整体的生存潜力因进化中每前进一步以及由之所形成的新体制而日渐增加？这个问题可以归纳为两种可能性：(1)通过经年累月的积加逐步使各个组分按顺序渐进式使体制在进化中得以改善，从而使每一组分与作为遗传器的其他组分相互作用成为一个功能实体之前，在发展过程中便已拥有本身的生存能力（独立进化）；(2)遗传器的起源从一开始就要求体现所有组分的相互依存（相依进化），也就是说，任何单个组分是没有价值的，如同在现存遗传器中所见到的实际情况那样（图 203）。

就理论角度来说，我们感到第一种提法比较容易理解，因为各种结构组分的排列顺序体制的进化仅仅需要：在进化过程中的某一时刻，一举产生一个新的具有结构功能特征的分子体制，而后随着时间的推移，再经历一系列类似的步骤，一个完整的原型遗传器

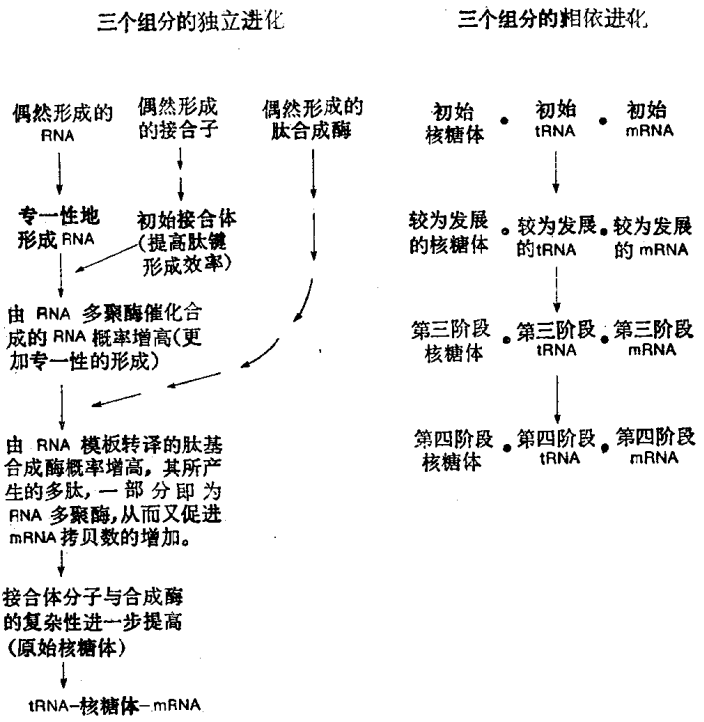


图 203　有关遗传器进化的两种相互对立的模型

旨在解释成为最大难题的有关遗传器进化方式的两种进化观点的图解：即tRNA，mRNA与核糖体三组分，分别发展，以后通过巧合进行结合(偶合)，从而形成现今的遗传器。对于tRNA类型的分子来说，或许早已借助选择准备好了。但是，更有可能的是，除了前生命浆液的条件之外，这类分子的进化是在有功效的原始的mRNA与肽键合成酶出现之后。图中所示的两种模型都有各自理论上的难题。不过对一个现实存在尚未得到解决的问题进行思维会有助于扩大知识，一般通过教学活动是不能得到这些知识的。

最终便能脱颖而出了。例如，各种各样的类tRNA分子(氨酰腺苷酸)可能在前生命环境中即已出现，因此作为最初的遗传器只要从这些已经准备完成的分子中挑选出几个佼佼者即可。

一般公认，在这一问题上的两种说法都会陷入理论上的困境，或许有朝一日会证明二者都不可靠，因之在其中进行取舍可能只不过是一种美学上的实践而已。但是不管怎么说，图204所提出的还是一个有关支持各组分结合成遗传器的独立进化模型较为合

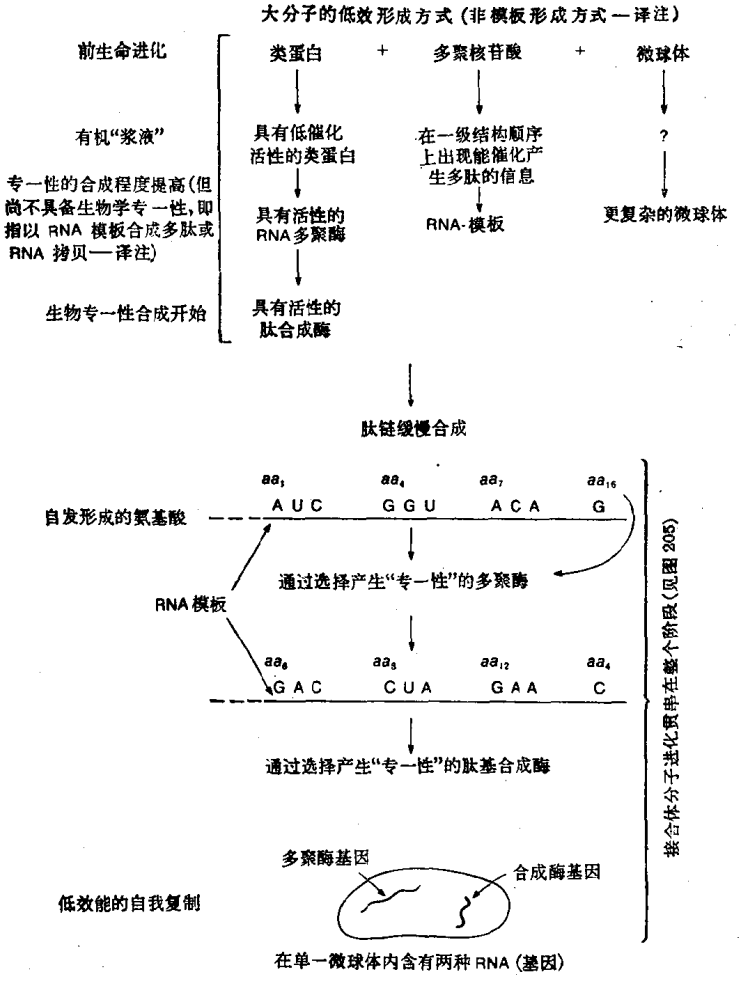


图 204　大分子物质系谱进化的管窥

在图中所绘的有关遗传器进化的模型中，读者会注意到在前生命进化末期到开始出现有效的自身复制系统之间存在着一段知识空白。一般来说，推动机体进化的因素都是在自身复制机理的基础上起作用的，本模型只要求在一个微环境内部只须形成某种"二个基因——二个酶"的系统，它将以远较前生命环境更为有效的方式产生各种大分子化合物。(在此系统中，肽基合成酶基因指导形成肽基合成酶，它将协助mRNA模板(多聚酶基因)产生RNA聚合酶，RNA 聚合酶会再使肽基合成酶基因和多聚酶基因增幅。最早的合成酶和多聚酶系由自然产生的类蛋白担任。最早的基因由自然选择产生。——译者注)。

理的设想。

读者在分析图 204 时最好要牢牢扣住生物学特性(指从RNA模板合成多肽的能力——译者)的进化以及结构功能的进化问题，这是因为特异性催化作用进化的问题人们常易把它视为进化的产物而予以忽视。现今的酶分子具有高度的特异性，能够促使反应速度增高数千倍之多。可是早先的类蛋白质分子其所起的作用恐怕也只略高于非催化反应的速度，至于反应特异性就更不用提及了。从发生一个反应本身来看，本来便无需要求什么高效性和高特异性，要使一个反应和反应系列能够免于湮没，全部的要求只是该反应发生的概率必须高于前生命时期可能发生的概率(这正是类蛋白可能起的作用——译者注)，一旦这类幸存的反应具有选择上的有利性，这时为使该反应进一步起动，唯一的附加条件只需要把它在一个微环境中固定下来，以免其选择效果为周围的大量有机浆液所冲淡。某些反应之是否能幸存与发展，全部得仰赖于一系列虽然轻微但却天天向上的发展过程，从而使之最终成为一个愈来愈加适应的系统。由此可见，反应的特异性也是逐步相应地增高的。同样，在进化过程中，特异性方向的变化也和结构上的复杂体制变化一样都必须要经受自然选择的检测。

其实，生物进化中所特有的相对特异性，在化学进化(即非生物进化)中也同样现实存在。在非生物进化中此类特异性甚至已达到这一程度：由前生命进化所产生的化合物分子并不表现为随机性组合，亦即事实上产生的化合物分子，只局限于其全部可能发生的分子结构组合中的一小部分范围。在 Miller-Urey-Fox 实验系统中产生的化合物分子也只是全部可能组合的极少数几种。其中有些类蛋白分子都具有微弱的催化活性，有些多核苷酸也无疑会在碱基排列上带有信息，要是能够转译，则势必会以更高的专一性指导合成多肽。不过，如前面所说的那样完美的核苷酸排列顺序以及氨基酸的排列顺序恐怕也未必可能在前生命进化阶段便已一气呵成。

生物进化是从有选择性的大分子集群，即很可能是从类蛋白

与多核苷酸分子组合开始发端的。在类蛋白分子中，某个（或某些）可能表现有 RNA 聚合酶的活性而与其他的类蛋白分子有所区别；在多核苷酸中，可能会有一个或多个分子的初级结构能够产生具有 RNA 聚合酶活性的多肽分子（如果能够转译的话）。问题是怎样来解释前生命聚合酶如何与前生命“信使 RNA”一起进而产生具有生物专一性（指由模板合成多肽——译者注）的多聚酶分子，以及从理论上来说明 mRNA 与多聚酶分子究竟又是如何与其他的类蛋白分子、多聚核苷酸划地为界的（以处于类细胞环境中与其他的化合物分子相互隔离）。如果某个或某些多聚核苷酸由于略微增高了合成酶的专一性而取得对前生命浆液中所必定存在的其他多数多聚核苷酸的选择优势的话，这里一个十分重要的因素就是因为该多核苷酸与多聚酶系处于一个封闭的类细胞结构中（在封闭的条件下，多聚酶有利于与之同存的多核苷酸大量拷贝复制，从而取得选择优势——译注）。

如果假定密码子为一个字母，一个字母正好对应于模板 RNA 中的一个核苷酸，并且每一个字母密码可以吸引数个氨基酸分子，这样我们便可能来说明有关编码的某些特异性了。RNA 模板至少由四种核苷酸组成（A、U、G 和 C），考虑到在前生命浆液中很可能存在有某些稀有碱基，因之组成 RNA 模板的核苷酸种类有可能更多。类蛋白质包括 20 种氨基酸（可能少一些）。结果形成的编码比率大体上是一种核苷酸约对应于 5 种氨基酸，这意味着每种核苷酸平均编码 5 种氨基酸（如果核苷酸种数大于 4 个或氨基酸种数少于 20，则该比率应相应降低）。

当然，应该指出，所提出的这种专一性程度很不充分，为了在漫长的发展过程中提供能专一性产生的多肽产物，亦即使之具有突出的能持续不断被有效合成的生存潜力，所有各种氨基酸很可能彼此之间对于碱基群（如上所说，对单个碱基只可能达到 5 对 1 的专一性，可能在对同一个碱基的 5 种氨基酸之间在亲和力上也有某些差异性——译注）的亲和性已经具有足够的差异。可是对初始接合子的进化来说，象这样一种氨基酸与碱基的对偶关系倒

不见得一定具有关键性性质(在 tRNA 分子中连接各种氨基酸的都是一种碱基,即腺嘌呤——译注)。

如图 204 所示,在提出这一模型时遇到的一个困难是关于解释氨基酸沿 RNA 模板结合到 RNA 模板上去的问题,这究竟是氨基酸与核苷酸直接的氨基酸-核苷酸方式呢,还是要通过一个接合子——中间媒介,二种说法都缺乏可供支持的证据。但是,一旦形成了一个氨基酸-RNA 模板复合物,要说明前生命类蛋白质在作为拟酶,识别氨基酸-RNA 模板复合物的基础上催化其相邻氨基酸形成肽键便不会过于为难了。该新形成的多肽和别的多肽的区别是其初级结构完全取决于模板 RNA 的初级结构,而且在合成方式上已经转变为酶促作用。

拥有 RNA 多聚酶活性的第二类多肽的存在是该模型中同样不可少的因素,这一类多肽能够识别作为底物的裸露模板 RNA,并使自由核苷酸据以合成为数众多的初始模板的复制品,这是前生命方式所望尘莫及的,换句话说,因为多聚酶有催化活性,结果必然会使多核苷酸(初始基因)的数目迅速增长,另一方面,这种多核苷酸的一级顺序又是为多聚酶"编码"的,于是多核苷酸行使两方面编码功能:(1)在转译多聚酶时作为 mRNA 起作用,(2)复制过程中又作为 RNA 模板。可以想象,复制过程也会涉及两个步骤,第一步先由原始的"+"链形成"-"链,第二步再由"-"链形成"+"链。处于这样原始系统中的各种化合物分子只要它们能够为类细胞环境所保护而免于稀释冲淡,只要它们的前体和肽基转移酶能保证由生命前浆液中得到源源供给,就会持续不断地扩增繁衍。

在转入讨论下一步进化过程之前,必须补充说明一下所想象的第一步进化的另一种可选择的方式。例如,还没有明显的理由把"活性"氨基酸的进化的起始阶段在第一进化阶段中避而不谈(图 205),现有的化学证据表明,在氨基酸与腺嘌呤之间有着一种自发的吸引力,能使氨基酸与腺苷酸彼此偶合起来,在现代接合体分子中关于氨酰-tRNA化合物的形成机理也正是用氨基酸与腺苷

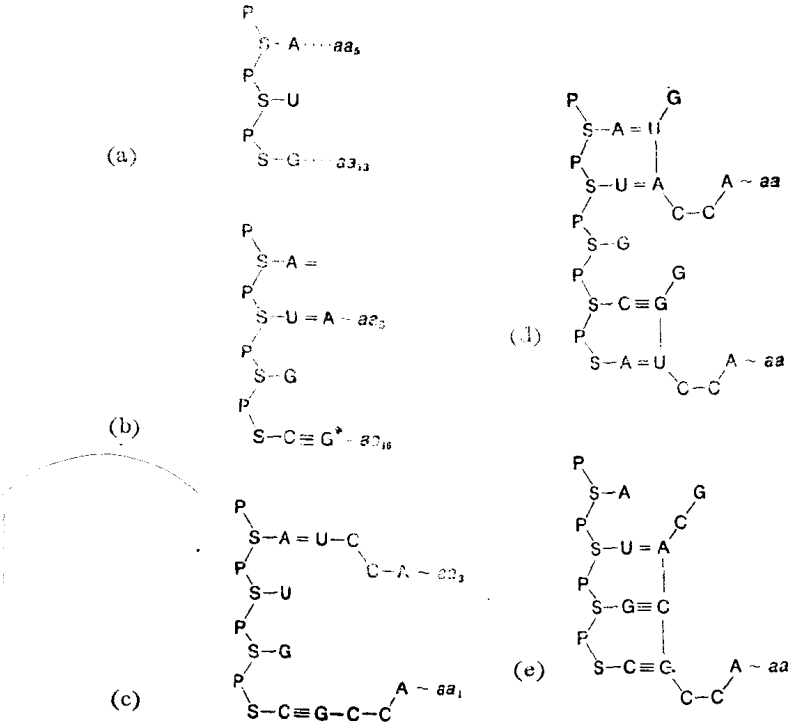


图 205 接合体分子的进化

(a) 如果将尚未澄清的有关信使核酸与氨基酸之间所显示的专一性归之于为氨基酸固有的属性，则第一个有关蛋白质合成的复合物便可能先于接合体分子而存在。

(b) 一个初始接合子可能就是一个(腺苷酸)或多个核苷酸。

(c) 接合体分子愈加复杂。

(d) 二字母密码接合子。

(e) 三字母密码接合子。

氨基酸的活化能与接合体分子大小无关，可能以化学方式(即由初始的类ATP 分子)提供，但更有可能是来自于热能、紫外线或电能。

* 原书的-G=G～系误——译注。

酸的偶合来解释的。但无论如何，现代接合体分子还携带有旨在使氨酰-tRNA 复合物识别 mRNA 分子的特异性，不过在最早的此类系统中，担负这一角色的则很可能是氨基酸。早先有一种看法，认为原始接合体分子通过选择获得的第一个功能就是通过供给能量和使氨基酸直线归位来促进肽链形成，这样看来，氨基酸的活

化作用即氨基酰基腺苷酸化合物的形成必定在不存在化学能源而只可能由热量、紫外线等提供能量的前生命环境中便已发生。可以想象，如果除腺苷酸之外，胞苷酸、尿苷酸、鸟苷酸同样能够作为结合子使氨基酸活化的话，接合子的此种由单个字母形成的密码都有可能会提供某种专一性(指与特定氨基酸偶合——译注)，目前关于多字母接合子的假说究竟是对是错都还没有什么可靠的实验给以作证。

假定一个初始的接合体分子就是一个腺苷酸，则接合子的进化就可以用腺苷酸变成多核苷酸不断加长延伸的方式进行解释。如果同意这一思想，便不妨进而假设在接合体分子进化过程中氨基酸直接对模板的专一性将会在某个时刻转变为接合子-模板专一性。在现代遗传体制中，氨基酰-tRNA 分子对 mRNA 分子的全部专一性都存在于氨基酰 tRNA 复合物中的 tRNA 一个部分，但既然某种体制型的专一性的接合子在起源上是否为蛋白质合成所绝对必需这一点尚不清楚，因此认为这样的解释较为顺理成章：即最早的专一性是来自于氨基酸与核苷酸的直接偶合，而接合子与核苷酸的专一性结合则是在以后进化中从前者转变而来的。

氨基酸的专一性转换为接合子专一性时，翻译过程中的编码关系也必须要随之转换，比如在转换之前，初始密码为单字母密码，亦即氨基酸在多肽链伸长期间定位时是以一个氨基酸随同一个单核苷酸接合，或完全单枪匹马地被吸引到模板分子单个核苷酸上的方式完成的，而在转换为接合子对模板 RNA 的专一性的场合，模板和接合子之间的编码关系便有可能采取单字母密码、双字母密码乃至三字母密码等多种形式，至于这三种编码方式，那一种最优，哪一种最劣，眼下还没有充分的理由加以评判。

从产生具有生物特异性的雏型 RNA 多聚酶来看，较之借助于前生命肽键合成酶作为催化剂更为有利的一种武器，是在于建立某种越来越依赖于由 RNA 模板指导的肽基转移酶才能得以成功运转的发展体制，这是因为这样产生的肽基转移酶远比前生命肽键合成酶更能专一性地形成和有效得多。为了能够发展这样一

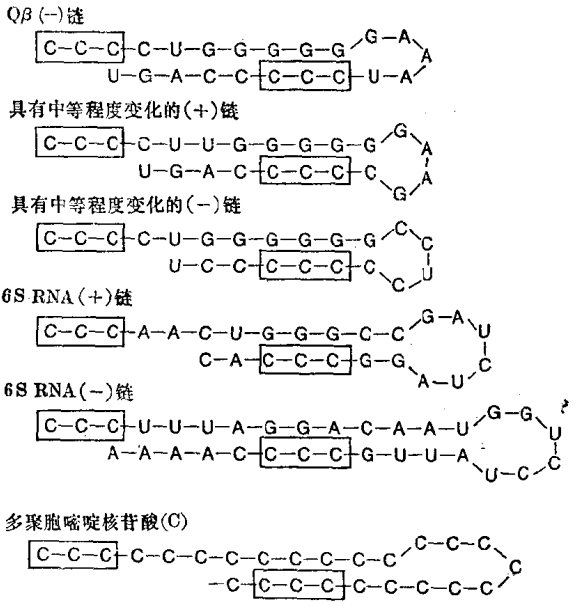
种体制，凡是有益于多聚酶存在和起作用的环境条件必须同样在某些方面也要有利于合成酶的存活和行使功能（否则二者便合不到一起——译注）。和初始的多聚酶一样，最早的合成酶都是类蛋白分子，催化活性低微，也就是说，从"生物学需要"来看，这类分子在持续存在保证供应方面是不可能确保的。而第一个具有"生物特异性"的合成酶却是一个与上述合成酶（即类蛋白分子）迥然不同的酶分子，其氨基酸的排列顺序则完全由 RNA 模板编码。因此，目的在说明这一体制的模型必须要能解释：通过二种模板和二种酶分子共处于一个能长久自我保存的微环境（细胞）中的偶合作用，结果使得系统的复制和繁殖终于取得比其周围的类蛋白及多聚核苷酸分子更大的优势。

关于两种 RNA 模板和由之编码的两种酶的共存形式合乎逻辑的解释是应该发生在一个微球体内部，因为由微球体提供的环境可以使得两种功能蛋白质在保证模板-酶复合物（即两种模板-两种酶的系统）迅猛增加繁衍方面取得成功的机率达到顶点。设若两个 RNA 分子碰巧以共价键结合起来，形成统一的带有两个"基因"的 RNA 分子，这样的"细胞"事实上便已包含有一条雏形染色体，将来能够产生两种酶。不过，就上述模型来说，早期阶段形成二个"基因"一条"染色体"的方式本身并不重要。实际正如下面实验所表明的，较小的 RNA 分子可能要比较大的分子复制得更快，如果说复制速度对复合体的存活具有举足轻重的意义，那么染色体的发生无疑将是在稍后一些的阶段。图 206a 所示即为有关 RNA 模板逐步取得选择优势的一个粗略轮廓。

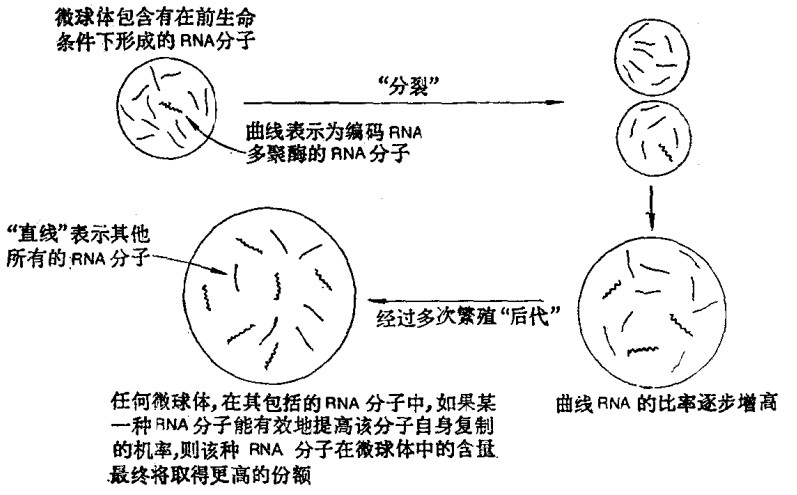
选择过程可能在 RNA 分子自我保存不被冲淡湮没方面起有作用这一点已经得到实验证明。Spiegelman 的实验是将 Qβ RNA 病毒与 RNA 复制酶一起在特定的环境中培养（在该环境中除了复制功能之外，Qβ 病毒的一切机能即使不能表现亦不致影响该病毒的生存），并使之处于一种严酷的旨在导致病毒 RNA 分子大小不断减缩的选择压力之下。其中复制较快的 RNA 分子由于选择作用而得以存留繁衍，终于逐步从复制较慢的 RNA 分子（由于严

酷的选择条件，复制较慢者不能完成复制周期的可能性亦高）中分离出来。当继续保持上述选择压力（即使全部已经是复制较快的RNA 分子，但只要 RNA 分子之间存在有复制速度的差异，选择便照样有效）的场合，所存留的 RNA 分子复制越快，其分子大小亦因之而越来越小，也就是说，在这样一种特定的环境条件下，分子复制可以摒弃掉一切在该条件下已属无足轻重的碱基顺序，而只要保留为 RNA 复制酶进行识别以及能顺利复制的核苷酸顺序即可。经过 74 代连续的转接培养之后发现，Qβ 病毒中有近 88% 的初始染色体业已被丢失，整个 RNA 分子总共只剩下 550 个核苷酸。

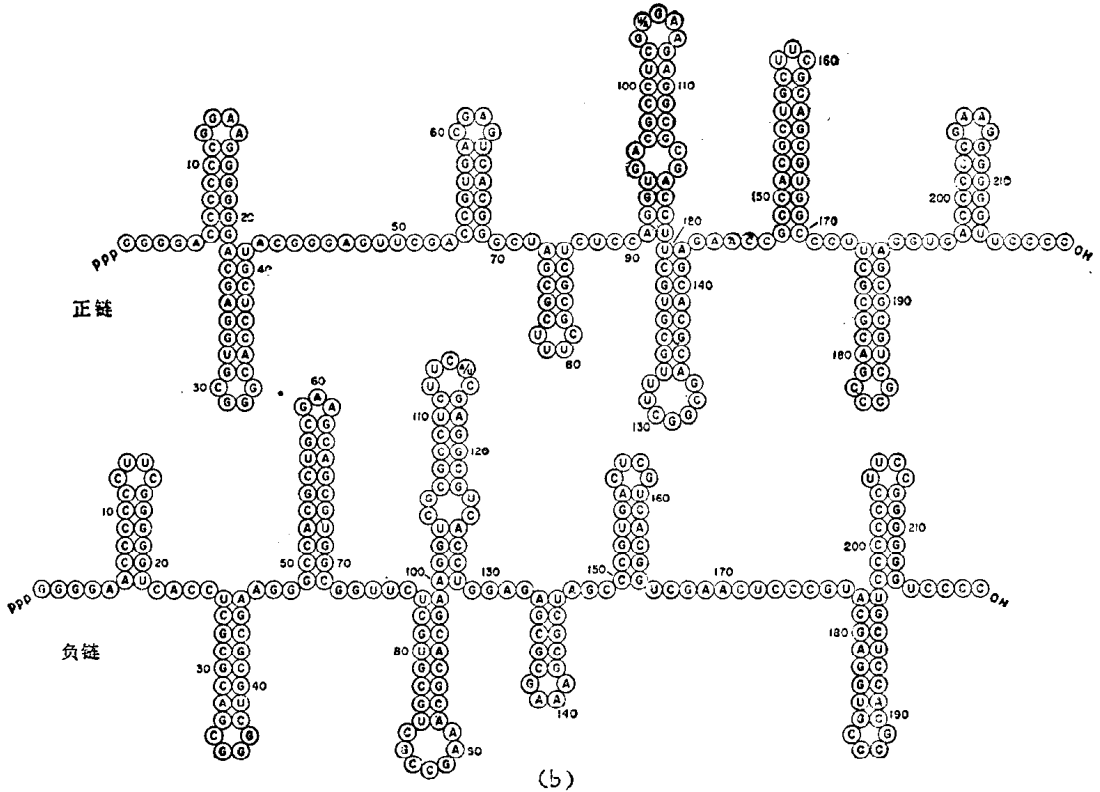
对于 Qβ RNA 分子复制来说，通过对选择压力作相应的调整和改变即可获得多种多样的小型 RNA 分子，其中的 MDV-1 只具有 218 个核苷酸单元，但其复制能力在一定程度上依然类似于完整的 Qβ 病毒。MDV-1 分子的碱基排列顺序已经确认，预计它同样存在有较高程度的次级结构，亦即在沿 RNA 单链上不同的区域，具有彼此互补的碱基配对后产生折叠（发夹结构），结果使整个 RNA 分子形成一些自身对合区域。在将 Qβ 病毒的（+）链与（−）链对照分析时，可以看到折叠的花样彼此互呈镜面像关系（镜影形，图 206b）。换句话说，有关被复制酶识别的信息，大概只存在于RNA分子的二级结构以及一级结构之中，而且（+）链和（−）链的信息预期完全相同，使用同样的选择方法，还选出另一个比 MDV-1 分子更小的 RNA 复制分子（微型 RNA），分离后检测，它含 114 个核苷酸，初级结构与 MDV-1 完全不同，因此，很可能 RNA 分子的初级结构并非为酶分子识别的唯一决定因素，其二级结构以及三级结构必定举足轻重，尽管人们对此尚无法作出解释。但是这并不意味着初级结构无关宏旨。在 Qβ 系统中，任何寡核苷酸或多核苷酸，只要 3' 端是 C-C-C 顺序，以及在 Qβ 复制酶所规定的空间位置上再有第二个 C-C-C 顺序，即可作为一个有效的模板顺序进行复制，例如，下列各种 RNA 分子均可作为复制酶的模板而行使其功能：



如果我们从 Qβ 系统来推测 RNA 分子初始系统的形成，那就有理由认为，演化初期对 RNA 分子的选择作用并非完全取决于该



(a)



(b)

RNA 分子所赖以置身的那个"细胞"的全部表型特征，而且也同样依赖于 RNA 分子本身的二级结构。后者作为能否保存而免于冲淡湮没的尺度来说，可能不会比复制速度以及为初始多聚酶所识别的效应有更多的内容。

核糖体是转译系统中最后一个需要说明的部件。如前所说，现今的核糖体是已含有多种蛋白质和至少 3 种 RNA 分子的细胞器的复合物。在核糖体蛋白质中有肽基合成酶以及其他许多因子——起译因子、伸长因子、终止因子。核糖体复合物的原始形式可能就是最早的那个既起合成酶的作用又起核糖体替身作用的合成酶。最初的合成酶的存在方式或许是作为某种比较大的RNA蛋白质复合物的一个组成部分，但是至今尚无法说明在这样一个复合物中是如何自我存留繁衍下来的。因此，最好把核糖体的演化解释为这样的一系例事件：即各种"协作蛋白"和 RNA 分子在完成自身进化的同时不断地组装到原先固有的合成酶上去。其中有些协助蛋白无疑是作为辅助因子发挥作用的，它们能启动蛋白合成酶，帮助多肽链伸长，并终止蛋白质的合成。

图 206　RNA 分子自我繁衍的选择模型

微球体内的一个模板RNA，因具有足够高的专一性催化产生"RNA多聚酶"的能力而得以选择生存，这是因为RNA多聚酶反过来又能提高新RNA分子形成的机率，新产生的RNA分子与微球体中原先存在的RNA分子如出一辙。

(a) 图中绘制的关于二基因的"细胞"（在这个"细胞"中，一个基因编码RNA多聚酶，另一个基因编码肽基合成酶）的进化模型（图中只是有关RNA多聚酶的基因部分——译注）并无实验证据，然而 Spiegelman 在 Qβ RNA 试验中提出的模型倒确实得到了某些实验的支持，如果RNA分子在某种特定的环境中（在这里只要保证复制酶识别的功能，即使其他机能全不表现亦足可使 Qβ 病毒存活）可以舍弃其他一切非属必须的机能而能得以存活，则选择作用将会有利于较小的RNA分子，迅速繁衍后代，这是因为较小的RNA分子复制时远较大的RNA分子为快。当然，RNA分子链短片的有利性应有其限度，即它至少必须保留有为复制酶识别的碱基顺序。

(b) 图示经多次连续转接后所选得的QβRNA分子正链（+）与负链（−）的碱基排列顺序。在转接期间，由于处于特定的环境之中从而使较小的RNA分子取得了选择优势。小型RNA分子的链长只及完整 Qβ 病毒的 5%。碱基顺序表明该链本身具有几个碱基配对区段（从初级结构推断）而且（+）链与（−）链的二级结构彼此雷同。（+）链从左向右看，（−）链从右向左看。（引自 D. R. Mills, F. R. Kramer 与 S. Spiegelman, Science, June 1, 1973。）

到现在为止，我们已讨论了两个 RNA“基因”，以及一个RNA多聚酶和一个肽基合成酶，它们存在于微球体内从而受到保护，人们不妨可以把如此一个不断在发展着的遗传器概括为某种拥有足够的遗传信息，以便在一个由前生命化学作用产生的多种化合物分子组成的生态系统中，保证得到生存下去的体制。从前生命的偶然形成信息转变为自身携带信息，这是一个渐变过程。通过突变和选择所带来的这一转变，其在进化上的意义就在于大大地提高了催化能力以及催化产物的专一性。图 206 便是就这一转变过程所作的图解说明，图中包含有曲线 RNA 的微球体，很可能就是那些含有更多机会存在的 RNA 多聚酶的微球体。如果由于另一种 RNA 分子的存在(最初当然是偶然性的存在)，如果由于此 RNA 能提高微球体可能含有肽基转移酶的机会，而能使上述那样的一种微球大为扩增，那末该微球体的生存潜力势必亦因而提高。由于微球体的生存潜力提高，故对于前生命“酶”的依赖性即相应的逐步减弱，只要这两种功能偶联起来，一个具有两基因-两酶的系统便会就此建立。有关遗传器的主要演化步骤可以归纳于下：

1．从初始的肽基合成酶向核糖体演化；

2．接合体分子专一性的演化与单字母密码向三联体密码的发展；

3．RNA 模板中数目增高单链增长，亦即基因在数目与复杂性方面都趋向于增加；

4．将信息由 RNA 传递到 DNA 的逆向转录酶发展之后，另一种多聚酶即担负着将信息由 DNA 传递到 RNA 的所谓依赖于 DNA 的 RNA 多聚酶亦行演化。

代谢途径的进化

即使有一些事实和资料的帮助，人们要重建进化的历史过程也是不容易的，尤其从遗传器来看，更没有多少可资利用的事实和资料，因而，这一任务更为艰巨。在上述讨论中还留有一些尾巴。对此，我们将把注意力着重放在依靠实验证据支持的有关生命进

化历史的三个主要发展阶段上：(1)代谢能力的强化，(2)基因组大小的扩增，(3)蛋白质一级结构的比较研究。上述进化史的三个阶段，每一步除了告诉我们有关生物信息流的演变、形成以及发展机制的若干情况外，同时还推测了一种通用的遗传密码以及与此相适应的遗传器。

1945年 Norman Horowitz 受 Oparin 启发率先对代谢途径的进化作了探讨，就如何解释由异养有机体进化到自养有机体的问题提出了自己的见解，其基本假定是：最早的有机体是从浩瀚无穷的"有机"分子进化而来的。在这样一种充满有机分子的环境中，只是为了延续"生命"的需要，有机体分子将逐渐组织、汇集成为大分子，并从小分子物质汲取能量。据信出现第一个完全异养的有机体乃是完全出于将含有能量的小分子合成为某种建筑部件的需要。

就如我们已经看到的那样，在早期进化阶段，前生命化学和生物化学的主要差别，在于生物的反应速度大大超过前生命的反应速度。在0～50°C这一狭窄的温度范围内尤其如此。鉴于这一点，对于一个靠聚合大分子化合物进行生长并从小分子化合物汲取能量的这样一类"细胞"来说，如果它们仅仅只仰赖于缓慢低效的前生命的活动方式来支撑自己，就会存在有把自身耗尽的危险。Horowitz 举精氨酸代谢为例指出，当精氨酸的前生命供给枯竭时，由于细胞对精氨酸的需求便会形成一种选择压力，从而有利于那些能够演化为某种新的方式攫取精氨酸的细胞生存。作者认为如果前生命浆液中有精氨酸存在，则完全有理由相信前生命浆液中也一定会包含有在分子结构以及化学特性上与精氨酸相近的某些分子，这样，有些细胞便能以最小的能量消耗，将此类化合物分子转化为精氨酸。在精氨酸不足所造成的压力之下，此类携带有能够催化类精氨酸分子成为精氨酸的多肽分子的细胞便会压倒那些缺乏这种能力的细胞而使自己存活下来。

将精氨琥珀酸催化成为精氨酸，所耗的能量不多。于是某些生活在前生命浆液中的细胞通过将精氨琥珀酸转变为精氨酸，便

开辟了一条供给精氨酸的新的来路。不过精氨琥珀酸到一定的时候也会同样耗尽。结果，细胞便不得不重演同样的过程以寻觅另外获得精氨酸的途径。譬如，能够将类西瓜氨酸分子转变为精氨琥珀酸，最后具有这一能力的细胞将又获得比其原先的细胞有更大的生存能力。最后当前生命浆液中连基本的有机分子也完全告罄时，那么就唯有那些拥有依靠酶催化而将无机分子变成为有机分子的代谢能力的细胞才能得以生存下去。这一论点是假定预先已有某种遗传器的存在，并借助于遗传、突变和选择作用，随之带动着遗传信息量的增加，其结果则是酶功能和代谢能力的提高。

基因组的扩增

现代物种中，基因组的扩增存在着多种方式，例如，一个基因组内的DNA能以全部染色体的重复来扩增(图207)，也能以一个或几个染色体成员甚至通过染色体的断片或几个碱基对的重复来扩增（图208、209)。DNA的扩增是现代物种的普遍现象，而对于绝大多数的情况来说，所有上列形式的基因组的重复，据知都与有性重组（表34）的过程有关系（虽然发生此种情况的频率极低)。

DNA的某种增加，仅就说明新变异来说就很有必要，这是因为它可以解释代谢能力的扩展进程。譬如，假如基因A对细胞的存活是必需的，那么基因A的突变就会导致细胞的死亡；这时如果基因A已经在染色体组内得到重复，则其中一个基因的突变就不致引起细胞丧生，因为未突变的基因*A*照样能执行为生存必需的基本功能。另一方面，虽然这些*A*基因中，某一个发生突变后，其直接效应往往不一定会有助于该细胞的生存，正如在“突变”一节中所指出的，产生无义、错义及中性突变的机会要比导致增加存活潜力的突变更多，但是当存在Horowitz假说所谓的选择压力时，在广泛范围的各种可能突变中会有一种特异性的突变却可能具有至关重要的作用。

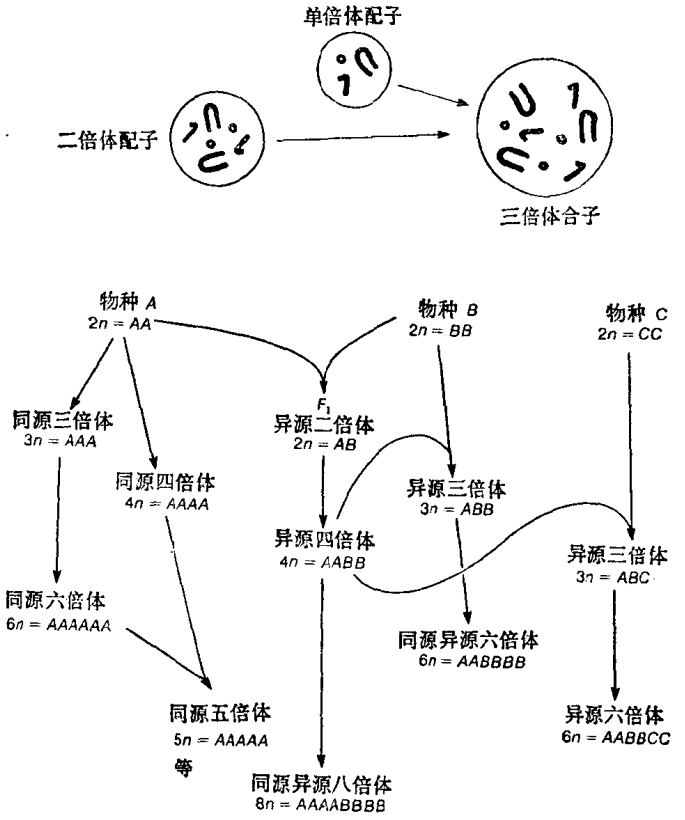


图 207　借助多倍体扩增基因组

图示染色体组扩增的各种形式，从三个二倍体品种 *A*,*B*,*C* 开始(分别用符号 *AA*,*BB*,*CC* 表示它们的二套染色体)，因而可能构成二倍体，三倍体，四倍体、五倍体、六倍体等等类型。同源多倍体这一术语指的是染色体组来源于同一物种，如 *AA* 为同源二倍体，*AAA* 为同源三倍体等等，异源多倍体这一术语指的是染色体组来源于不同的物种，例如，*AB* 为异源二倍体，*AABB* 为异源四倍体等等。在自然群体中，异源二倍体通常不育，因为在减数分裂时，*A* 组染色体无法和 B 组染色体正常配对。如果能将异源二倍体加倍而成为异源四倍体，那么由于两套 *A* 染色体及两套 *B* 染色体恢复正常配对从而成为可育。至于三倍体植物，因为在减数分裂时同源染色体不能两两配对，故也是不育的，但许多三倍体植物均可借助无性繁殖而繁衍不绝(如苹果树)。(引自 Stebbins, 1950，及 Lewis 与 John, 1963。)

如果两组 *A* 基因作用产生两组 A 蛋白，则载体细胞就可具有生产比满足生存所需数量多得多的 A 蛋白质的潜力。此时设若诸如通过置换一对碱基的方式使一个 *A* 基因发生突变，于是便会产

生出另一种蛋白质A′。新蛋白A′同A蛋白之间在化学性质上差别可能不大，但在功能上则可能很不相同。但是再怎么说，A′蛋白比之那些偶然形成的新的蛋白质分子来，总更接近于A蛋白。如果A蛋白的分子能够识别几种底物的话，也许A′蛋白至少也能识别

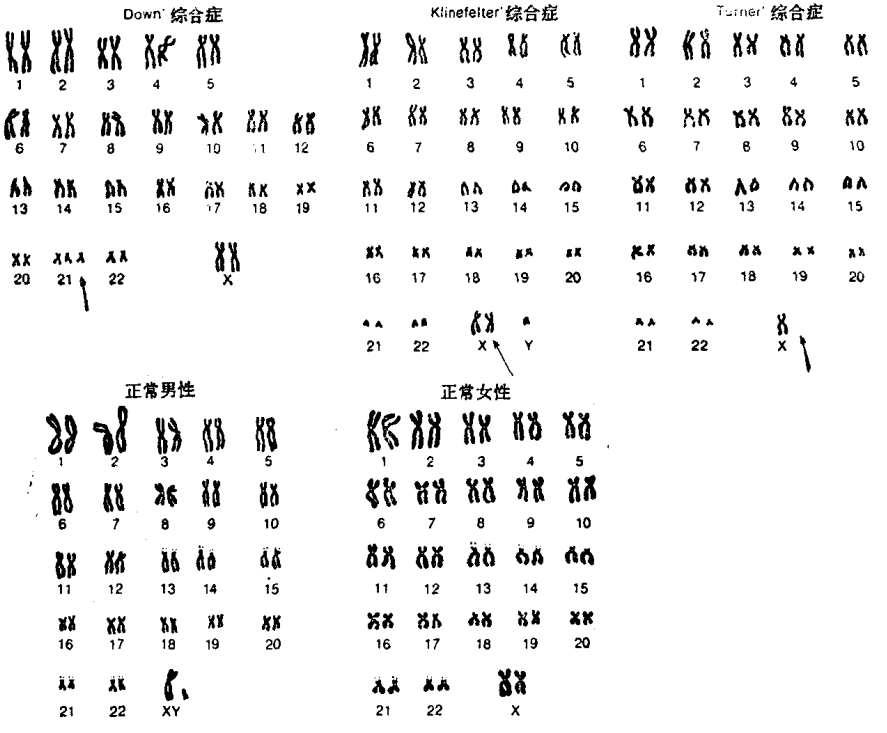
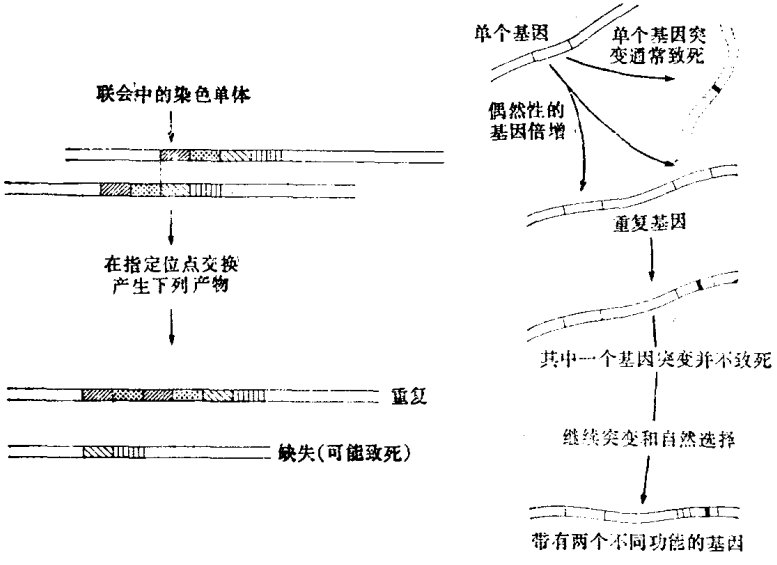


图 208　人体染色体组通过增加一条或多条染色体而扩增

男性或女性的全部染色体数目均为46条，在这23对染色体中，有一对染色体男女间彼此不同，这便是性染色体，性染色体具有X与Y两种类型，男性只有一条X染色体，女性有两条X染色体。Down氏综合症与Klinefelter氏综合症患者都是比正常个体多一条染色体，(Down氏综合症病员具有三条第21染色体，Kleinfelter综合症病员具有2条X染色体和1条Y染色体)。Turner氏综合症患者则与之相反，而是缺少一条Y染色体。在减数分裂中，如果发生一对同源染色体中的两个成员同时进入一极，亦即同时进入一个将来会发育为配子的一个子细胞时，即会招致上述各种染色体数目的异常表现。当一个具有22条常染色体加两条X染色体的卵细胞与另一个正常的包含有一条X染色体或一条Y染色体的精子结合，其形成的后代要末是一个具有XXX的女性，要末便是一个XXY的Klinefelter氏综合症的患者。(引自 Strickberger, 1976。)

性	染色体	表现特征
女性	XO(单体)	Turner 综合症
	XX(二体)	正常
	XXX(三体),	过度女性症
	$XXXX$(四体)	过度女性症
	$XXXXX$(五体)	过度女性症
男性	XY(二体)	正常
	XYY(三体)	身高超过正常男性
	XXY(三体)	Klinefelter 综合症
	$XXYY$(双二体)	Klinefelter 综合症
	$XXXY$(四体)	极度 Klinefelter 综合症
	$XXXXY$(五体)	极度 Klinefelter 综合症



(a)

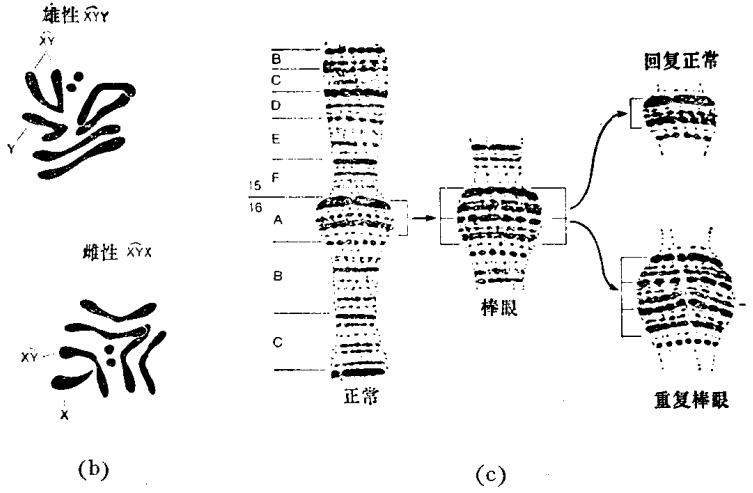


图 209　通过染色体片断重复来扩增基因组

引起染色体内部结构的改变，表现有各种方式，其中有些是：

(a) 不等价交换：左边表示两条同源染色体之间发生错配，已经错配的同源染色体之间发生一次交换，导致减数分裂的子代中其中一个子代染色体上很小的一段重复，而另一子代染色体的相应部位则表现出缺失。右边图说明的一种方式是重复基因的一个突变能够产生一个新的功能基因，同时又保留了原有基因的功能。(引自 Dickerson 与 Geis, 1969。)

(b) 诸如用辐射等方法使染色体发生断裂时，这时断裂端往往具有"粘性"，因之在用辐射处理后，染色体会照例发生重新排列。本图说明了果蝇中 X-Y 染色体之间发生粘连的情况：一类是全套染色体组加一条Y染色体，另一条则是附加一条X染色体。使用光学显微镜，我们一般无法看到所发生的许多结构变化，比如DNA 整合到质粒上(图 72)以及碱基对的增加和缺失等等(图 91)。

(c) 图中所示为果蝇中唾腺巨型染色体上棒眼区段的重复。(引自 Calvin Bridges, 1936。)

其中之一种。再从精氨酸代谢来看，如果本来的 A 蛋白既能识别精氨酸又能认别精氨琥珀酸，突变后的 A′ 蛋白虽然不能认别精氨酸了，但也许还能识别精氨琥珀酸。如果这样，那末对 A′ 蛋白来说，在其上再发展出一个瓜氨酸识别位点，亦即使精氨琥珀酸识别位点和瓜氨酸识别位点两者并存的机会，肯定比之一个与 A 蛋白毫无关系的蛋白质能够使两者同时偶尔发育并存的机会要大得多(图 210)。

如同在沙门氏菌 (*Salmonella typhimurium*) 中观察到的组

表 34　由基因扩增而产生的彼此较为类似的蛋白质分子：展示结构趋异的程度从而带来功能上的差别

蛋　白　质	差异程度(百分数)	分化时间估计(百万年前)
血红蛋白与肌红蛋白	77	1100
乳清蛋白和溶菌酶	63	350
血红蛋白 α 与 β 链	59	600
胰凝乳蛋白酶 A 与 B	21	270
血红蛋白 β 与胎儿血红蛋白(偶蹄动物)	19	40
血红蛋白 β 和 δ 链(人类)	8	40

(引自 M. O. Dayhoff, 1972。)

氨酸基因一样，控制着有关组氨酸合成代谢路线上各个酶的基因都彼此紧紧地串联在一起（图 188），这种结合方式同样也适合于代谢上相互关联的其他一些基因。功能上彼此密切有关的基因在染色体上连成一串这一事实虽然还不是一个证据（指有关基因的起源——译注)；但是基因这种串集方式却可以用由不等价交换之后产生的基因重复以及紧接着的基因突变来给予解释(图209)。目前还不大可能就重复、不等价交换、异倍性以及多倍性等等在基因组不断增大的进化过程中所起的作用作一确切的判断，虽然这些现象是客观存在的事实。在进化阶梯上处于较高阶段的物种一般来说都要比较原始的种属拥有更大的基因组(表 35)。

分析比较现存物种的系统发育关系有多种方法，其中之一就是比较它们之间基因组的大小。为了深入探讨这个问题，可以根据一个细胞进行生命活动所需要的酶的数目来估算它所具有的基因数目。大肠杆菌的基因数估计为 4×10^3左右，这个数目相当于 4×10^6个碱基对，由此而推算的染色体大小接近于大肠杆菌的染色体长度。作为对比，哺乳动物的每个基因组 DNA 平均大约含 3×10^9个碱基对，这个容量用于制造 3×10^6个基因是绰有余裕的。根据极为粗略的估算，维持哺乳动物生命所需要的基因总数甚至

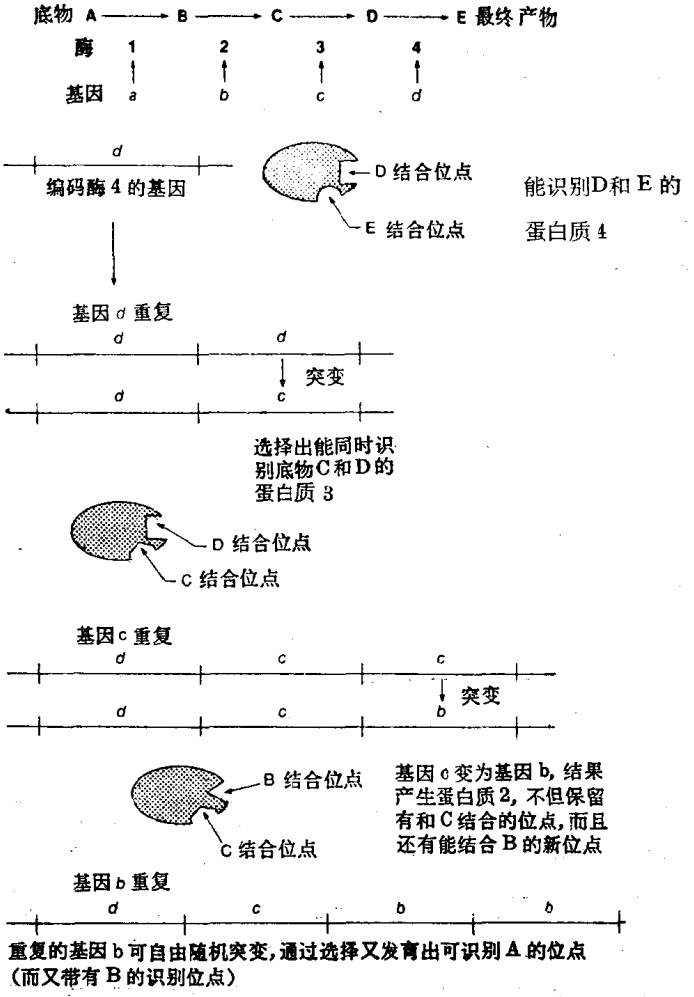


图 210　酶分子进化假想机制

在这个假设的代谢途径中，蛋白质 4 有着识别前体D以及末端产物 E 的位点，它并能催化由D到E的转换反应。在前生命条件下，如果酶蛋白 4 是为原始有机体所携带的属于这代谢途径的最早一个酶，也是仅有的一个酶的话，那么该原始有机体就能继续生存直到前体D在环境中成为限制因素为止，正是在这时，选择压力将有助于蛋白 3 出现，它能催化另一些化合物分子（前体 C）来形成前体D，前体C分子也是早在前生命时期的环境条件中就已经存在的。从理论上讲，这个过程可以一直持续到最终产物E能利用简单的无机分子以及碳源来合成为止。由图可见，在每一种情形下，已加倍的基因均可通过突变来饰改成对基因中的一个基因，而又不丧失整个原有基因的功能。此外，由于这两个蛋白质在任何情况下都还得识别一个共同底物（对一个来说是作为前体，对另一个来说是作为产物），因而为新的催化功能所需要的饰改可降到最低限度。

表 35　基因组大小

	整个基因组平均值 (10^{-12}克 DNA)	单倍体的核苷酸数目 ($\times 10^9$碱基对)
哺乳类	6.5	3.2
鸟　类	2.4	1.2
爬行类:		
海龟	5.0	2.4
鳄鱼	5.0	2.4
蜥蜴	3.8	1.9
蛇	3.8	1.9
两栖类:		
蛙及蟾蜍	12.8	6.2
美西螈	9.3	4.5
其他	45～181	21.8～87.9
真骨鱼类:		
大部分种类	1.8	0.9
鳟鱼	5.2	2.5
鲤鱼、金鱼	3.2	1.6
弓鳍鱼	2.3	1.1
鲟鱼	4.3	2.1
腔棘鱼	5.6	2.7
肺鱼(两个种类)	100, 230	48.5, 111.7
鲨　鱼	5.9	2.9
无颌类	2.5, 5.1	1.2, 2.5
原索动物:		
文昌鱼	1.1	0.5
被囊类动物	0.3	0.2
棘皮动物	1.7	0.8
节肢动物:		
甲壳纲	3.0	1.5
昆　虫:		
蝗蜂	11.7	5.7
果蝇	0.2	0.1
软体动物:		
鱿鱼	9.0	4.4
其他	1.2	0.6
环节动物	2.9	1.4
腔肠动物	0.7	0.3

（续上表）

	整个基因组平均值（10^{-12}克 DNA）	单倍体的核苷酸数目（$\times 10^9$碱基对）
海　绵	0.12	0.06
原生动物：		
纤毛虫	14～1,670	
孢子虫	0.59	
变形虫	1,400	
鞭毛虫	2.4	
藻类	0.1～200	
真菌（子囊菌类）	0.064	0.031
细菌	0.013	0.013
动物病毒	$0.3\sim42\times10^{-5}$	$0.3\sim41\times10^{-5}$
噬菌体	$0.2\sim59\times10^{-5}$	$0.2\sim57\times10^{-5}$

（引自 M. O. Dayhoff,1972。）

不及这个数目的四分之一。至于病毒和细菌，一般来说，它们的基因组则似乎是按维持基本生命活动所必需的基因数目所规定着的大小剪裁过的。随着物种的结构和功能逐步复杂化，基因组也就变得越来越大，甚至超过能够根据设想所计算的功能基因数目，那么这些所谓“过量”DNA 究竟具有何种功能呢？它们又为什么会超过如此之多呢？

一般情况是过量 DNA 的数量随着系统发育的复杂化而增加，但高等真核类（脊椎动物）中却出现普遍的例外，比如两栖类中的 *Amphiuma*，每个基因组有 10^{11}碱基对，原始的肺鱼大约是 5×10^{10}个碱基对，而其他与之有关的种类和许多哺乳类每个基因组差不多只有 10^9碱基对，为什么功能需求相对“较少”的两栖类，其每个基因组含有的 DNA 反而会比哺乳动物更多呢？

几年前，发现了一类“新”的 DNA，它同信息 DNA 在两个主要方面存在着差别：其一，它是否参与遗传信息流的传递尚无定论；其二，它是高度重复的，并且这些重复的碱基顺序同结构基因相比相当之短（只有 6～15 个核苷酸对）。在这类高度重复的 DNA 区段中，已经知道有的以二个碱基为一重复单位，在 DNA 长链中

进行多次重复(如蟹的多聚 dAT 即可连续重复达 50 次之多)。从功能上看,这样一些重复顺序的作用似乎还不大清楚。

在前面讨论为 rRNA 编码的基因时,曾经指出从核仁形成区分离得到的 rDNA 要比其余的 DNA 具有更高的密度(GC 含量较多),起初大家相信,所有的重复顺序大概都可以按密度把它们与其他的非重复顺序 DNA 区分开来(这些 DNA 与一般的 DNA 不同,因而在密度梯度中表现出其特有的分带花样,称为随体 DNA),但是现在已经知道,所谓的高密度 DNA 仅仅不过是整个各种类型的重复 DNA 的一部分而已。

运用 DNA 分子杂交技术是估计一个基因组中重复 DNA 数量的最有用的办法。双链 DNA 变性后,两条互补链能够通过密度梯度离心技术拆开,但当再次混和时,两条原已分开的重链和轻链会经所谓退火处理而重新恢复成双螺旋结构。在进行上述变性-退火实验时,往往把其中一条链(或轻链或重链)切成一些小的片断,而使另一链保持完整(保持完整的这一条链在实验中一般保证充分过量),并令其固定在滤纸或凝胶上,然后再放上 DNA 单链片断,这样 DNA 片断就会在凝胶上自由泳动,在完整链上"寻找"自己的互补部分。如果一个顺反子在基因组只出现一次,那么它找到互补链并很快退火重组的可能性自然要比该顺反子具有较多重复的场合可能性要小一些,倘若一个顺反子有上百万次的重复,则所形成的双链部分的数目相当可观,退火-重组的过程亦将迅速完成,由此可见,测定"退火"速度的快慢乃是推算重复 DNA 数量的可靠线索和可信赖的手段。

在小鼠基因组中,估计约有 70% 的 DNA 是独一无二的无重复顺序(即每个基因组中一个顺反子),退火速度亦数此类DNA顺序最慢,但一旦退火,它就具有最为精确的特异性。另外大约20%的 DNA 则表现重复,每个基因组的重复顺序数目在 10^2～10^5之间,这类 DNA 退火速度比那些无重复顺序的 DNA 要快,但由于每个基因组包含的各个重复顺序的数目,即拷贝数涨落很大,因此在退火速度方面也同样会参差不齐。还有剩下的 10% DNA 在密

度上与前二类 DNA 不同，退火速度最为迅速，其重复顺序每个基因组为 2×10^7个，并且具有别具一格的碱基成分(GC含量占 34%，而在无重复的以及中度重复 DNA 中 GC 含量为 42%)。此类高度重复的DNA每个重复单位的长度通常在 5～15 个核苷酸对数，大部分集中在染色体的着丝点及核仁形成区内。

DNA 重复顺序的类型多种多样，且因物种不同而相迥异，但是，不管何种类型，这些重复顺序总是沿染色体的长度相挨排列的。一般说来，来自亲缘关系较为密切的物种的重复 DNA 之间不会发生碱基的交叉配对，这意味着物种的亲缘关系即便相近，其所携有的重复 DNA 顺序亦未必相同。每个基因组所含有的高度重复 DNA 的总量一般变化于 1%到 30%之间，视不同生物种属而异。有的物种其所含的重复顺序的类型只有一种，而另一些物种譬如果蝇，重复顺序的类型甚至可多达 5 种。在豚鼠中曾发现一种高重复 DNA，长度是 6 个核苷酸对，退火速度极快。对这个短片的核苷酸排列顺序测定后得知，轻链为 5′-CCCTAA-3′，重链为 3′-GGGATT-5′。在豚鼠中还有另一些 DNA 重复顺序，有些重复顺序似乎是上述重复顺序型式中的某些碱基作某种转换，有一些则又表现为某些碱基的颠换。

高重复的 DNA 是不是以同无重复的单一顺序的 DNA 同样的方式进行转录迄今尚无确证，关于重复顺序的功能究竟如何也提不出任何材料。但是这一切有可能暗示我们，高重复 DNA 除了提供某种不完备的 DNA 材料外，并不具一般 DNA 作为模板转录的功能，而新的机能则很可能便是从这些 DNA 材料中衍生出来的。按照这一思路推想，每个基因组都可能备有为新的基因与新的酶分子所预定的 DNA 保留区。或许这一合乎理性的设想有其真理的内核，但它却无法解释为什么有些生物会采用这一方式"预定前景"而别的生物却又并非如此。

高重复 DNA 很可能是在染色体 (DNA) 配对时发挥功能作用的，这是人们对其含义的另一种推测，例如，重复 DNA 区段的缺失似乎并不影响机体的生活能力，可是减数分裂中染色体不乐

离的频率却会确有增加，结果也降低了育性。又譬如，从重复 DNA 大多局限于染色体上异染色质部位来看，则暗示着它的作用与同源染色体联会可能有关。这一推想曾获得某些实验观察的支持：小鼠的 Y 染色体是独一无二的，其随体DNA 顺序并不严格限定位于着丝粒左右部位，另一方面，它又是 40 条染色体中唯一不能有效配对的一条染色体。根据重复 DNA 在染色体上部位局限化的事实所获得的另一个启示是特别在染色体生活周史期间，重复 DNA 可能有助于维持染色体结构的稳定性。

尽管人们尚未掌握有关重复 DNA 功能的过硬证据，但众所周知，在所有哺乳类动物的基因组内均包含有高重复 DNA，另外，将小鼠细胞在某种 DNA 易于丧失的条件下进行培养时，随体 DNA 仍然保持完好不见有丧失迹象。如果从这一角度来看，即细胞在进化过程中根本无法承担过度的能量消耗，那么根据上述的实验启示以及一般的逻辑推理，可以认为随体 DNA 之所以不离不弃一定是因为：重复 DNA 本身对于细胞一定是不可欠缺的组成部分。

极少数几种中等重复的 DNA 业经鉴定具有特殊的功能，rDNA 即为其中之一。在生物体内有一小类基因，它们经转录后并不进一步合成特异性蛋白质，rDNA 也属于这一类基因。在细菌中，一般未发现有高度重复的 DNA 顺序，但是 rDNA 基因居然会重复 40 次之多。至于在一些比较复杂的生物中，每个基因组中rDNA 基因重复 100～1000 次者亦不足为奇，目前正在获得的证据表明，rDNA 基因在更高级的物种中一般甚为相似，甚至在系统发育上判若云泥的物种之间也不相上下。

另有一类专为组蛋白编码的重复 DNA 顺序则极端保守，不易发生变化，在海胆中此种 DNA 在每套单倍性的基因组中重复可达 1000 次，人的 HeLa 细胞中为免疫球蛋白可变部分编码的高保守的重复基因，重复亦有 300 次之多。在某些生物中，为 5S RNA 及 tRNA 编码的这类 DNA 基因，其复重顺序的排列很可能是前后相互串联的。另一方面，高重复 DNA 似乎又有着对进化

颇为重要的高周转速率(turnover rate)，有事实为证：一般进化上最先进的生物种属，重复 DNA 最多。而在进化上亲缘相近的种属却常常并不具有彼此共通的重复顺序。确实，在分析了豚鼠的随体 DNA 之后才恍然大悟，碱基置换是这样普遍，以致于由碱基代换造成的 DNA 结构变化太快而无法依靠自然选择的力量使之稳定下来。

David Hogness 对中等重复的 DNA 所作的最有意义的观察是有关两个质粒杂交的实验，此两质粒携带有明显可供识别的 *Drosophila*(果蝇)重复 DNA 的区段。质粒分子杂交情况表明，在果蝇染色体区段，在许多相互间隔的位点上都能发生碱基配对，位点分散有可能启示此种中等重复 DNA 在功能上或许拥有某种调节作用。除了 Hogness 的工作之外，也还有其他一些证据，譬如中等重复 DNA 在海胆和爪蛙(*Xenopus*)的染色体中，其 DNA 顺序与众不同，亦呈散置状态。中等重复 DNA 的这种零散分布性质可能更适宜于协同有致地对基因群实现控调，这里所说的基因群并不如 *lac* 操纵子基因群那样是串联集结在一起的。

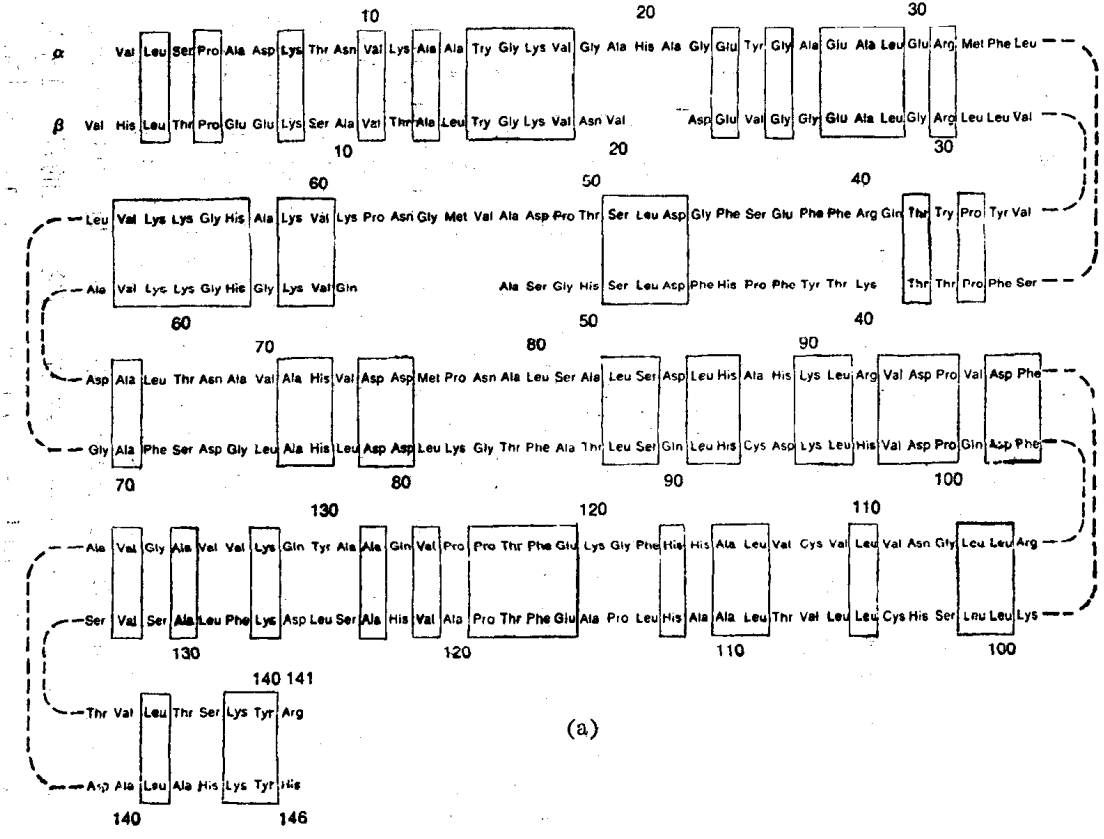
如果说在进化进程中，一些 DNA 是稳定保守的，另一些 DNA 是高度可变的，那么要问变化速率究竟是受什么控制的呢？一切 DNA 分子的结构和组成都极为相似（指以磷酸去氧核糖构成的骨架以及作为结构单元的 4 种核苷酸都完全一样——译注），这一点使得原先认为某些 DNA 可能比其他 DNA 更易于突变的假说很难立得住脚，事实上，在一个物种内部，高重复的可变 DNA 与高保守(不变的) DNA 之间周转速率的差别要比不同物种信息 DNA 分子之间的差别还要大些，关于这些问题的线索可参见蛋白质进化一节。

蛋白质进化

当自然选择对物种以及物种中的各个个体起作用时，它究竟是直接作用于 DNA(即基因)呢，还是作用于基因产物(RNA或蛋白质)，亦即作用于表现型？从根本来说，全部表现型乃是决定机

体本身能否存活下去的决定因素。诚然，按照达尔文的说法，所谓进化是指整个物种的动态现象，并不是什么生物个体遭遇的命运，从这一意义出发，人们可以把自然选择看作为主要是对基因型施加压力的因素：它通过时间作为试金石来对各种基因型进行比较检测，并据以来判定每一个基因型的存活价值。因而现在仿照自然选择评判基因型优势的方式来理解基因型（全套基因）的发展进程，将会较深刻地把握进化现象的本质。例如，有些基因因为由之决定的表现型有利于在生存竞争中存活而得到保存并成为物种基因组中的添加成分，这便是进化；反之，如果某些基因决定的表型不适于生存而从群体中淘汰无存，这同样也表现为是一种进化。然而，如果把进化仅仅理解为某些基因的增加和消失，则又将问题简单化了，事实上进化过程是十分复杂的综合性事件，这一点可以从下面的讨论中得到说明。如有些基因渊源久远，但在整个进化的历史时期变化细微，乃至它们于现存的所有物种都大同小异，即使在最简单与最复杂的种类之间，也只有极少数几个核苷酸的差别。这意味着，自然选择的力量既可以造成生物表型上的天渊之别，但与此同时又使如此五光十色的生物种属保持着某种十分一致的功能（实质上也就是保留了同样的基因）。对此下面拟略述数端。

从包括血红蛋白和肌红蛋白（如前所述）在内的球蛋白家族可以为揭示作为进化中的一个侧面的保守性提供一点论据。在这里，球蛋白是一个大分子化合物家族，其特点是具有共同的载氧功能（氧分子载体），但初级结构变化很大（图 211），肌红蛋白从肽链的长度看在这个家族中可以说是一个与众不同的成员，因为比之于别的球蛋白分子它要多出几个氨基酸残基。球蛋白分子一般可分为三类，肌红蛋白算一类，人类和马血红蛋白 α 链归在第二类，所有血红蛋白的其余 β、γ 以及 δ 链构成第三类。从图 211 我们可以看到，就肽链内部氨基酸成员的差别程度而论，人类的 β、γ 链差异最小（仅 10 个氨基酸不同），马和人的 α 链次之（有 18 个残基不一样），如此等等。借助于这种种比较，就有可能勾画出如同反映物种之间解剖学亲缘关系那样的所谓球蛋白谱系树（图 212）。



(a)

肌红蛋白		Gln	Asp	Ilu	Leu	Ilu	Arg	Leu	Phe	Lys	Ser	His	Pro	Glu	Thr	Leu	Glu	Lys	Phe	Asp	Arg	Phe	Lys	His	Leu	Lys	Thr
血红蛋白	马 α	Ala	Glu	Ala	Leu	Glu	Arg	Met	Phe	Leu	Gly	Phe	Pro	Thr	Thr	Lys	Thr	Tyr	Phe	Pro	His	Phe	....	Asp	Leu	Ser	His
	马 β	Gly	Glu	Ala	Leu	Gly	Arg	Leu	Leu	Val	Val	Tyr	Pro	Trp	Thr	Gln	Arg	Phe	Phe	Asp	Ser	Phe	Gly	Asp	Leu	Ser	Gly
	人 α	Ala	Glu	Ala	Leu	Glu	Arg	Met	Phe	Leu	Ser	Phe	Pro	Thr	Thr	Lys	Thr	Tyr	Phe	Pro	His	Phe	....	Asp	Leu	Ser	His
	人 β	Gly	Glu	Ala	Leu	Gly	Arg	Leu	Leu	Val	Val	Tyr	Pro	Trp	Thr	Gln	Arg	Phe	Phe	Glu	Ser	Phe	Gly	Asp	Leu	Ser	Thr
	人 γ	Gly	Glu	Thr	Leu	Gly	Arg	Leu	Leu	Val	Val	Tyr	Pro	Trp	Thr	Gln	Arg	Phe	Phe	Asp	Ser	Phe	Gly	Asn	Leu	Ser	Ser
	人 δ	Gly	Glu	Ala	Leu	Gly	Arg	Leu	Leu	Val	Val	Tyr	Pro	Trp	Thr	Gln	Arg	Phe	Phe	Glu	Ser	Phe	Gly	Asp	Leu	Ser	Ser

肌红蛋白		Glu	Ala	Glu	Met	Lys	Ala	Ser	Glu	Asp	Leu	Lys	Lys	His	Gly	Val	Thr	Val	Leu	Thr	Ala	Leu	Gly	Ala	Ilu	Leu	Lys
血红蛋白	马 α	....	....	....	....	....	Gly	Ser	Ala	Gln	Val	Lys	Ala	His	Gly	Lys	Lys	Val	Ala	Asp	Gly	Leu	Thr	Leu	Ala	Val	Gly
	马 β	Pro	Asp	Ala	Val	Met	Gly	Asn	Pro	Lys	Val	Lys	Ala	His	Gly	Lys	Lys	Val	Leu	His	Ser	Phe	Gly	Glu	Gly	Val	His
	人 α	....	....	....	....	....	Gly	Ser	Ala	Gln	Val	Lys	Gly	His	Gly	Lys	Lys	Val	Ala	Asp	Ala	Leu	Thr	Asn	Ala	Val	Ala
	人 β	Pro	Asp	Ala	Val	Met	Gly	Asn	Pro	Lys	Val	Lys	Ala	His	Gly	Lys	Lys	Val	Leu	Gly	Ala	Phe	Ser	Asp	Gly	Leu	Ala
	人 γ	Ala	Ser	Ala	Ilu	Met	Gly	Asn	Pro	Lys	Val	Lys	Ala	His	Gly	Lys	Lys	Val	Leu	Thr	Ser	Leu	Gly	Asp	Ala	Ilu	Lys
	人 δ	Pro	Asp	Ala	Val	Met	Gly	Asn	Pro	Lys	Val	Lys	Ala	His	Gly	Lys	Lys	Val	Leu	Gly	Ala	Phe	Ser	Asp	Gly	Leu	Ala

肌红蛋白		Lys	Lys	Gly	His	His	Glu	Ala	Glu	Leu	Lys	Pro	Leu	Ala	Gln	Ser	His	Ala	Thr	Lys	His	Lys	Ilu	Pro	Ilu	Lys	Tyr
血红蛋白	马 α	His	Leu	Asp	Asp	Leu	Pro	Gly	Ala	Leu	Ser	Asp	Leu	Ser	Asn	Leu	His	Ala	His	Lys	Leu	Arg	Val	Asp	Pro	Val	Asn
	马 β	His	Leu	Asp	Asn	Leu	Lys	Gly	Thr	Phe	Ala	Ala	Leu	Ser	Glu	Leu	His	Cys	Asp	Lys	Leu	His	Val	Asp	Pro	Glu	Asn
	人 α	His	Val	Asp	Asp	Met	Pro	Asn	Ala	Leu	Ser	Ala	Leu	Ser	Asp	Leu	His	Ala	His	Lys	Leu	Arg	Val	Asp	Pro	Val	Asn
	人 β	His	Leu	Asp	Asn	Leu	Lys	Gly	Thr	Phe	Ala	Thr	Leu	Ser	Glu	Leu	His	Cys	Asp	Lys	Leu	His	Val	Asp	Pro	Glu	Asn
	人 γ	His	Leu	Asp	Asp	Leu	Lys	Gly	Thr	Phe	Ala	Gln	Leu	Ser	Glu	Leu	His	Cys	Asp	Lys	Leu	His	Val	Asp	Pro	Glu	Asn
	人 δ	His	Leu	Asp	Asn	Leu	Lys	Gly	Thr	Phe	Ser	Gln	Leu	Ser	Glu	Leu	His	Cys	Asp	Lys	Leu	His	Val	Asp	Pro	Glu	Asn

(b)(1)

肌红蛋白		Leu	Glu	Phe	Ilu	Ser	Glu	Ala	Ilu	Ilu	His	Val	Leu	His	Ser	Arg	His	Pro	Gly	Asn	Phe	Gly	Ala	Asp	Ala	Gln	Gly
血红蛋白	马 α	Phe	Lys	Leu	Leu	Ser	His	Cys	Leu	Leu	Ser	Thr	Leu	Ala	Val	His	Leu	Pro	Asn	Asp	Phe	Thr	Pro	Ala	Val	His	Ala
	马 β	Phe	Arg	Leu	Leu	Gly	Asn	Val	Leu	Ala	Leu	Val	Val	Ala	Arg	His	Phe	Gly	Lys	Asp	Phe	Thr	Pro	Glu	Leu	Gln	Ala
	人 α	Phe	Lys	Leu	Leu	Ser	His	Cys	Leu	Leu	Val	Thr	Leu	Ala	Ala	His	Leu	Pro	Ala	Glu	Phe	Thr	Pro	Ala	Val	His	Ala
	人 β	Phe	Arg	Leu	Leu	Gly	Asn	Val	Leu	Val	Cys	Val	Leu	Ala	His	His	Phe	Gly	Lys	Glu	Phe	Thr	Pro	Pro	Val	Gln	Ala
	人 γ	Phe	Lys	Leu	Leu	Gly	Asn	Val	Leu	Val	Thr	Val	Leu	Ala	Ilu	His	Phe	Gly	Lys	Glu	Phe	Thr	Pro	Glu	Val	Gln	Ala
	人 δ	Phe	Arg	Leu	Leu	Gly	Asn	Val	Leu	Val	Cys	Val	Leu	Ala	Arg	Asn	Phe	Gly	Lys	Glu	Phe	Thr	Pro	Gln	Met	Gln	Ala

肌红蛋白		Ala	Met	Asn	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Lys	Asp	Ilu	Ala	Ala	Lys	Tyr	Lys	Glu	Leu	Gly	Tyr	Gln	Gly
血红蛋白	马 α	Ser	Leu	Asp	Lys	Phe	Leu	Ser	Ser	Val	Ser	Thr	Val	Leu	Thr	Ser	Lys	Tyr	Arg						
	马 β	Ser	Tyr	Gln	Lys	Val	Val	Ala	Gly	Val	Ala	Asn	Ala	Leu	Ala	His	Lys	Tyr	His						
	人 α	Ser	Leu	Asp	Lys	Phe	Leu	Ala	Ser	Val	Ser	Thr	Val	Leu	Thr	Ser	Lys	Tyr	Arg						
	人 β	Ala	Tyr	Gln	Lys	Val	Val	Ala	Gly	Val	Ala	Asn	Ala	Leu	Ala	His	Lys	Tyr	His						
	人 γ	Ser	Trp	Gln	Lys	Met	Val	Thr	Gly	Val	Ala	Ser	Ala	Leu	Ser	Ser	Arg	Tyr	His						
	人 δ	Ala	Tyr	Gln	Lys	Val	Val	Ala	Gly	Val	Ala	Asn	Ala	Leu	Ala	His	Lys	Tyr	His						

(b)(2)

图 211　血红蛋白 α、β 链的氨基酸顺序

(a) 多肽链的氨基酸残基从N端起数，图中系将α、β链并列展示。两链的片断中凡带有相同残基的均以方框标示，不同的残基不打方框标示。α-链的 141 个残基中有 61 个与β链完全相同，即使在互不相同的残基位置，也总可发现功能相似的常见氨基酸，如在 6,7 位点上为天冬，谷，62,57-丙，甘，111,106-缬，亮，104,99-精，赖，72,67-丝，苏。在将两链对称配齐时需要作一些前后移动，但即使如此，两条链之间的一致性也是极其明显的。

(b) 肌红蛋白同几条血红蛋白链氨基酸顺序的比较。从这些球蛋白分子的N端开始，前后对准，以便显示其最大的顺序一致性，可看到它们之间明显的系统发育关系。图中空缺处(用破折号代替氨基酸)是表示一个蛋白链顺序中缺失的遗传信息，或是表示与之相应另一个蛋白链顺序中相对增加了一个遗传信息。(引自 M. O. Dayhoff, 1972。)

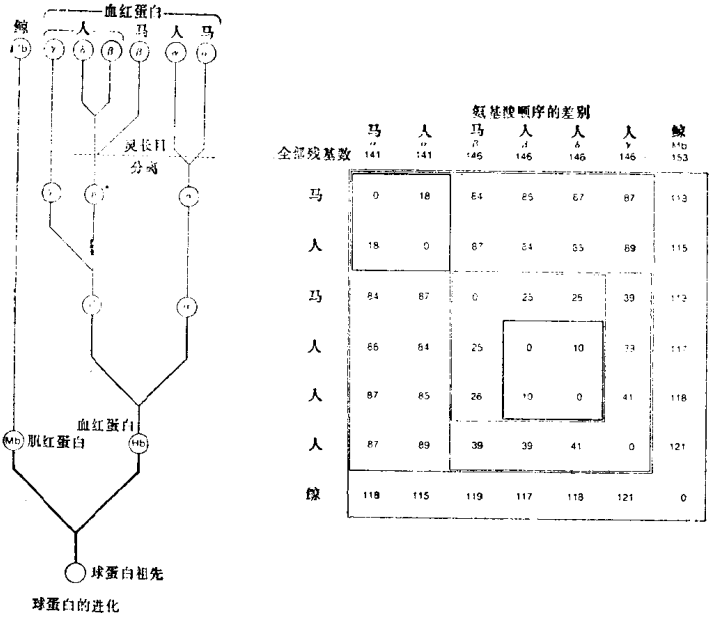


图 212　根据血红蛋白初级结构绘制的系谱图

本系统发育树是在比较血红蛋白链初级结构的基础上绘成的。由图可见，人类的β链和δ链最为相近，马与人的α链次之，肌红蛋白同所有的血红蛋白分子相比在差别程度上都大同小异。这个谱系树暗示着，鲸、马以及人类可能存在有一个共同的祖先，其球蛋白链有单一的氧结合位点，推想这条链很可能与肌红蛋白相类似，在以后的某一时刻，最初的那个基因必定产生了重复和突变，并最终形成某些更适合于氧分子输运的球蛋白分子。（引自 Dikerson 和 Geis, 1969。）

如果我们把视线从球蛋白分子的一级结构转移到球蛋白的三级结构，便会找到更多有关球蛋白家族亲缘关系的证据。（读者在图 213 中见到的是存在于人生命周期不同阶段的各种球蛋白含量的相对变化情形。）在血红蛋白中，α 链和 β 链差别最明显的部位是在D螺旋区段，这里 α 链虽少了构成D螺旋区的 5 个残基（从 D_2 到 D_6 共 5 个残基），但是，并不妨碍血红蛋白分子其余部分残基的折叠构象。另一方面，血红蛋白和肌红蛋白分子的三级结构则有着惊人的相似之处（图 161），这可能是某些不变位点（即各种球蛋白分子对应位置上的相同氨基酸）维持血红蛋白和肌红蛋白之间共同的三级结构起有重要的作用。

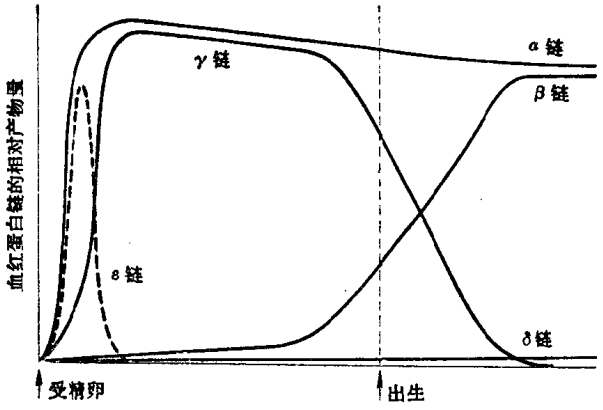


图 213　人体中各种球蛋白链的先后发育顺序以及人体 α 链和肌红蛋白链三级结构的比较

最早在人体胚胎中出现的球蛋白是α链和ε链，以后α链继续存留，ε链则在2～3个月内消失。在ε链消失的同时，γ链即开始形成。婴儿出生后不久γ链解体，代之而起的则是β链。在人体的整个个体发育周期δ链始终存在，但浓度极低。球蛋白分子的初级结构可能互不相同，但它们的三级结构却表现了明显的相似性(图161)。α链中有26个残基在一切类型血红蛋白分子中都恒定不变，其中有11个残基与肌红蛋白分子的相应区段也完全一样。(引自 Zuckerkandl, 1965。)

通过比较细胞色素 *c* 的一级结构，我们将会看到一幅更加完整的物种关系图。细胞色素 *c* 是类似于球蛋白的含有血红素的蛋白质，共具有 104 个氨基酸残基(分子量为 12400 道尔顿)，它与球蛋白不同之处仅仅在于细胞色素 *c* 的多肽链中，以共价键同血红素基因相连接的是两个半胱氨酸残基。

细胞色素 *c* 在功能上亦有别于球蛋白，它与其说是一种氧载体，毋宁说是一种电子载体，它所包含的铁原子是经其他参与电子传递的酶而被氧化或还原的。由于细胞色素 *c* 是所有细胞能量消耗和贮存系统的一个组成部分，因而它要比球蛋白分子更广泛地分布于现存的生物类群之中。在所有仰赖于好氧代谢的细胞类型中，从出现好氧代谢过程时起，细胞色素 *c* 或其功能类似物就可能已经同时存在了。可是对于球蛋白分子来说，它应是动物中原来的氧分子的扩散机制满足不了对氧分子输运要求时才应运而

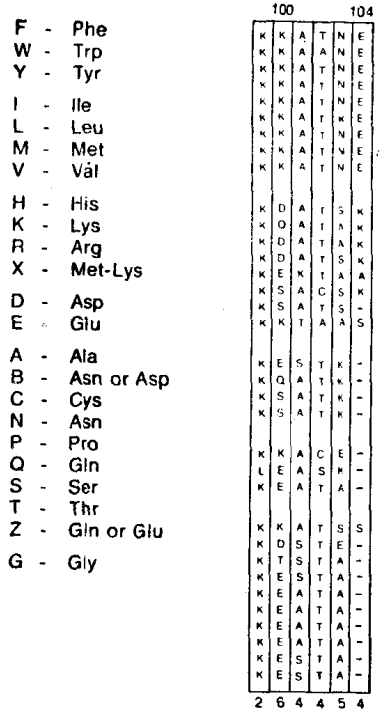
生的。

进化趋异性的问题一般能够借助于比较现代物种中功能相似的有关蛋白质一级结构的异同来加以估算，将这种估算作为考察物种之间历史关系的测温计有些场合十分可靠，有的场合则未必尽然。在实践上，一般由于无法从历史上对某种蛋白质结构进行追索，因之往往都是利用处于不同进化阶梯上的现存物种有关蛋白质分子之间的差别来提供认识其历史面貌的借鉴。迄今为止，有关这方面研究最有说服力的资料是各种生物细胞色素 c 一级结构的比较。经过科学家们艰苦卓绝的研究，现已基本搞清了分布于从酵母菌开始直到万物之灵人类的广大生物界的 34 种生物细胞色素 c 分子的氨基酸顺序(图 214)，结果发现，其中有 1/3 以上的氨基酸残基在分属于 34 个物种的细胞色素 c 分子中是共同的，这些氨基酸残基在分子中所占的位置称不变区，其中包括了从第 70 到第80 个残基，以及联接血红素的半胱 14，17 和组氨酸18 这几个最为突出的残基(图 215)。与不变区相对应的是变异区，细胞色素 c 的变异区因不同物种而异，由于它们属于三级结构范畴，因而很难从一级结构水平上对它们作出精确的描述。具备了这些知识，我们就能正式对几种细胞色素 c 分子的一级结构和立体结构作比较了。

利用 X 光衍射分析技术，可以把细胞色素 c 分子的功能部位在三度空间中确定下来 (图 216)。譬如血红素基因的位置正好嵌在一个豁口里边，靠共价键与豁口的二壁相连。槽的一边扣住血红素的是半胱 14，17 与组氨酸 18，槽的另一边似乎是由甲硫 80 与血红素相连。比较处于氧化态和还原态的马心细胞色素 c 的结构，我们就会发现，随着铁离子由 +3 价变成 +2 价的氧化态势(本身还原，氧化对方)时，整个分子的三级结构也会跟着改变。

我们还可通过图象比较来观察残基不变区与残基可变区的结构变化。细胞色素 c 分子中，直接和血红素相接的残基如第 70～80 这 11 个氨基酸是不变的，马心细胞色素 c 有 12 个甘氨酸残基，它们都有着高保守的趋向，另外，在某个由 8 个苯丙氨酸和酪氨酸

高等植物
低等植物
昆虫
其他脊椎动物
哺乳类
不同的氨基酸数目
a—有甲基基团
h—无甲基基团



①人②狒猴③马④驴⑤牛、猪、羊⑥狗⑦兔⑧加里福尼亚灰鲸⑨大灰袋鼠⑩小鸡、火鸡⑪鸽子⑫北京鸭⑬海龟⑭响尾蛇⑮牛蛙⑯金枪鱼⑰狗鱼⑱小柏天蚕⑲烟草天蛾⑳螺旋(锥)蝇㉑果蝇㉒面包酵母㉓克鲁斯假丝酵母㉔粗糙链孢霉㉕小麦胚㉖荞麦种子㉗向日葵种子㉘蚕豆㉙花椰菜㉚南瓜㉛芝麻㉜蓖麻籽㉝棉籽㉞蔺麻籽

图 214　34 种动植物材料细胞色素 *c* 的初级结构

细胞色素 *c* 在所有的真核类生物中都担负有同样的功能，其初级结构亦极为相似，在分子中某些区段的氨基酸顺序恒定不变，但其余区段则显示有相当的变化。不同种属细胞色素 *c* 的一级结构互有差异，但三级结构大体一致。(引自 R. E. Dickerson, 1972。)

组成的区段中有 7 个残基恒定不定或者只表现为二者之间的互换(即苯丙氨酸换成酪氨酸或相反)。这里尤其值得一提的是苯丙氨酸，它是不变氨基酸，由于它在被氧化还原时会发生结构上的迷离变幻而引起科学家们的极大兴趣：当处于氧化态势时，尽管苯丙氨酸本身为疏水性分子表面又包有水膜，它却总是暴露在细胞色素 *c* 分子的外部；而处在还原态势(本身氧化，还原对方)时，苯丙氨酸就会松弛而遁入分子内部。苯丙氨酸在氧化还原态势中所呈现的这种位置上的变换既暗示着它可能有某种功能作用，另外，这本身也为该苯丙氨酸在进化过程中的不变地位提供了合理解释。

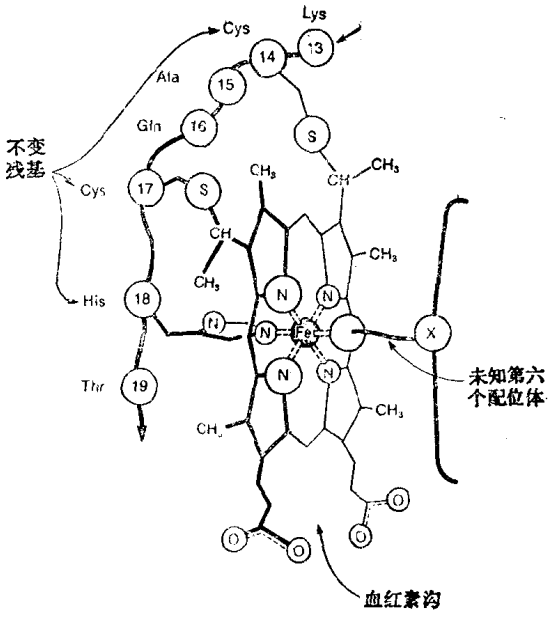


图 215　细胞色素 *c* 分子中的几个不变区

细胞色素*c*分子中的血红素基团位于与该分子表面成直角的血红素槽中，通过其左侧的半胱-14，17 及组-18 以及可能还有右侧的甲硫-80 固定在槽壁上。这些氨基酸都属于不变残基成员。(引自 K. E. Dickerson, 1972。)

在细胞色素 *c* 多肽链的可变区，不同种属来源的分子其相应位置上的氨基酸残基互不相同，但是一般地说这些相互替换的氨基酸之间仍然有一些相似的化学特性，例如，亲水氨基酸残基为另一亲水氨基酸残基所替代，疏水氨基酸残基为另一疏水氨基酸所替代，中性以及其他类型的氨基酸亦同样如此，等等。然而，亦偶或发生这种情况，比如第 89 个残基，会发生基团的改变，变化的范围可从碱性到酸性，乃至包括极性残基、微疏水性的不带电荷残基、带侧链的残基、和不带侧链的残基。但是在该位置上始终未发现过它能被一种大疏水残基所置换，细胞色素 *c* 的主要保守区域系位于分子"左"侧以及血红素所置身的袋(指豁口两壁)中，"右"侧是易变区(对于每个残基位置来说，氨基酸种类的变化范围在 1 个、2 个、到 3 个之间)。而"背"侧是分子的最易变区(每个位置上

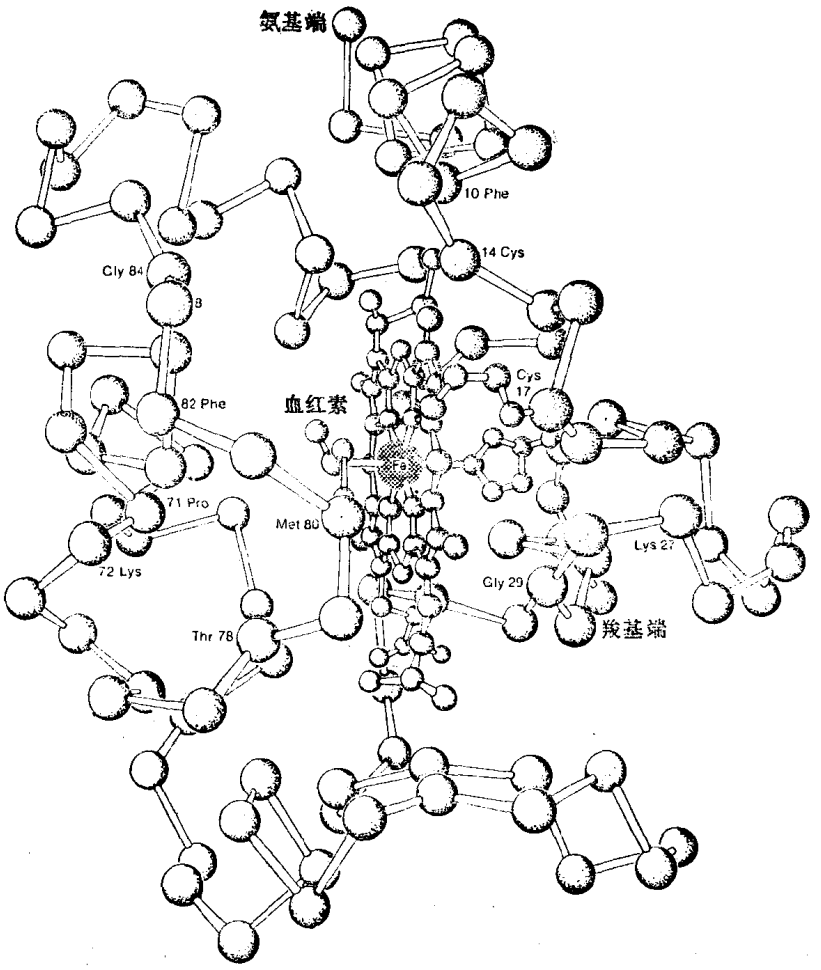


图 216　细胞色素 *c* 的三级结构

这个细胞色素*c*的三级结构图具有三维空间的特征，离读者愈远残基画得愈小。图中，从氨基端到羧基端可以追循大部分肽链中的残基方向，并可能在三级结构内部据以确定血红素基团的位置。(引自 R. E. Dickerson, 1972。)

残基种类的变化在 6 个以上)。

疏水残基存在于细胞色素 *c* 分子表面这一事实告诉我们，维持这种状态同样需要能量，而且当这些疏水残基在来到表面或者离开表面时需要肽链作某种折叠，这是不言而喻的，下面需要来解

释一下不变区残基的保守性。肽链真实的折叠过程我们并没有直接观察到，但是富于想象力和合乎逻辑的猜测和分析往往可以有助于我们弄清过程的真相，当苯丙氨酸这个疏水残基从分子外表转入内部或者反过来由内部返回到外表面时，由于整条肽链必须旋转方能保证苯丙 82 达到上文叙述的还原态势或氧化态势所在的位置，因而可以想见，在肽链旋转过程中分子内部一定包含有某种极有意义的构象变化。

细胞色素 c 同其他少数几个已用同样方法研究过它们的一级和三级结构的球蛋白分子之间的区别还在于，它的赖氨酸残基不仅位于分子外部，而且集结成由一个负电荷区隔裂开来的两个正电荷区(图 2.17)，由于所有的细胞色素 c 分子均有这种独特的排列方式，因此可推知，它对于细胞色素 c 分子发挥自己的正常功能是必需的。科学家们一直认为，细胞色素 c 与细胞色素氧化酶之间的吸引力是由于两者正、负电荷区产生静电引力的结果，而且细胞色素 c 的两个赖氨酸正电荷区中只有一个参加作用，这一推想得到实验观察的支持：位于"右"边正电荷区附近的赖氨酸当用化学方法使之变成三硝基苯赖氨酸时就会使细胞色素氧化酶失活，这可能是因为它阻断了细胞色素 c 同细胞色素氧化酶之间静电结合的缘故。至于细胞色素 c "左"侧的赖氨酸区(此区段赖氨酸较多，亦即另一正电荷区——译注)则可能是与细胞色素还原酶的结合部位。

细胞色素 c 分子中最易变的部分集中在两个正电荷区（即富赖氨酸区）之间的负电荷区，共包括 6 个残基位置。虽然从作为分隔两个正电荷区（即分别能同氧化酶还原酶结合的两个富赖氨酸区）的隔离区来说，其功能并非是不重要的，即便它是那末默默无闻，但易变的事实表明，它是整个分子最缺乏功能活性的区域。

在刚开始研究细胞色素 c 的结构时，由于还不知道甘氨酸残基具有一种高度专一的化学特性可借以分析推论整个分子的三级结构与功能表现，因此很不理解为什么甘氨酸在细胞色素 c 分子中会是一个恒定不变的残基，可是后来当有人观察到有些多肽链

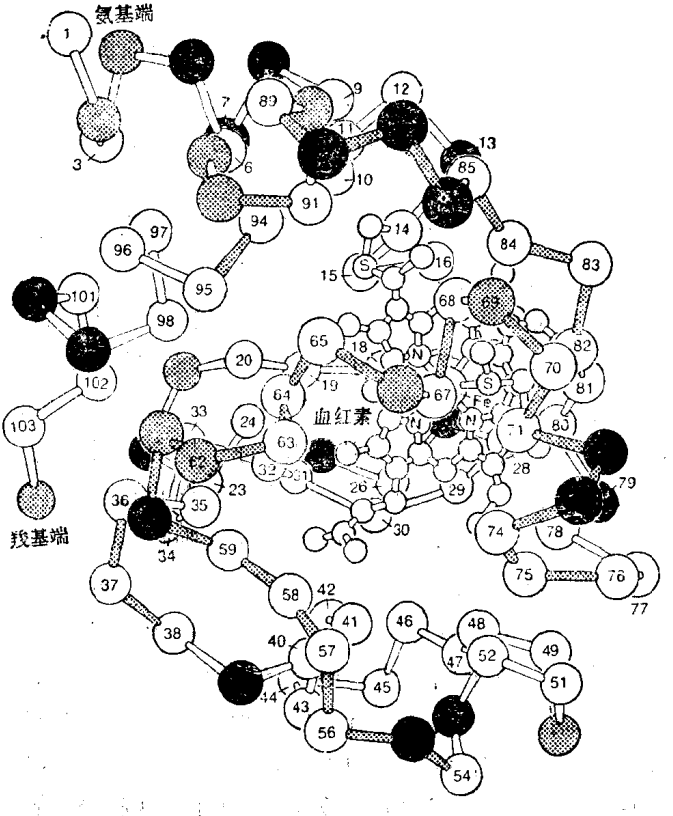


图 217　细胞色素 *c* 分子的带电残基

细胞色素 *c* 分子同球蛋白分子是有区别的，它的 19 个赖氨酸（黑色者）残基中，大部分发现于分子表面，带负电荷的 12 个残基（灰色者）中有 9 个也是在分子表面。带负电荷的残基集中在一起这一点意味着负电荷区或是参与一个细胞色素分子同别的细胞色素相连接，或是当电子由一个细胞色素流向另一个细胞色素时起着电子通道的作用，甚至二者兼而有之。（引自 R. E. Dickerson, 1972。）

片断是那么紧紧地包裹着血红素基团时这一谜底方才揭开，因为肽链这一部分的氨基酸残基裹装状态是如此致密，以致于很显然，惟有分子最小的甘氨酸才是最合适的材料。

芳香氨基酸的不变性也不容易解释清楚，它们在蛋白链自身折叠的区域成对出现只不过是它们在氨基酸顺序中外表上比较容易觅到的形式而已，譬如，第 10 与第 97 个芳香氨基酸的出现在豁

槽"右"边，第 46、48 个在豁槽"下"方，第 67、74 个则在疏水性的槽"左"侧，色氨酸 59 也在近旁，两个酪氨酸 46、48 通过与血红素基团的丙酸侧链结成氢键有助于血红素定位(图 217)。豁槽"左"方的三个芳香氨基酸可能在还原过程中参与电子传递，显然苯丙氨酸 82 也是肯定参加这一过程的。在还原态势时，血红素沟为多肽链中含有苯丙氨酸 82 的那一部分所关闭，关闭的方式是：苯丙氨酸 82 向"右"移动，填入血红素沟的"左"侧，这样，血红素基团就不大可能会受到外面的影响。可是，即使有了这些线索，为什么正是这一类型的结构体制对还原态势细胞色素 *c* 分子保持其生物学活性是那么重要这一点仍然不大清楚。看来，苯丙氨酸 82 很可能参与了活性部位调节，或是起着保护这个活性部位的作用。

当读者能从我们上面述说的角度去看待蛋白质的初级结构和高级结构时，就会对遗传与变异的种种问题形成新的见解。作为一种不太严密的讲法，进化可以被看作是一种时间现象，它或者淘汰或者保存了偶然产生的突变，从这一角度看，这意味着突变既是进化的原始材料，也可能是进化的限制因素。人们发现，在生物界的某些领域(珊瑚)，进化速度相当缓慢，而在另外一些领域(具有高重复 DNA) 变异则又极其迅速，那么能不能说控制着这种变化速率的就是 DNA 的突变性呢？从细胞色素 *c* 的讨论，回答是否定的，恰恰相反，变化速率的决定因素乃是出自于蛋白质结构和功能本身的某种抑制因素，比方说，在整个动植物界专为苯丙氨酸 82 编码的密码子并不是一个可靠的能保证苯丙 82 始终不变的抑制因素，该密码子本身犹如其他任何密码一样是可以突变的。但是，倘若一个突变导致一个其他什么氨基酸代替了第 82 位原应为苯丙氨酸的位置，便会使蛋白质的功能削弱甚至丧失，结果也就降低了生物的生活能力。

在今天要想根据蛋白质结构的知识来推测它们的功能并不容易，同样，要想在一般了解蛋白质功能的基础上来设想各种氨基酸残基在进化过程中的周转速度，其困难程度并不亚于前者。然而，象细胞色素 *c* 这样"普遍"地持续不断地起作用的蛋白质，比那些

仅存在于为数不多的几种生物中并只在短暂时间内起作用的蛋白质，要具有更高的保守性。从现有比较研究过的各种蛋白质看，初级结构上不存在广泛变异的蛋白质肽链(即表现为保守性)实在凤毛麟角，恰恰相反，变异乃是普遍现象，例如血纤维蛋白肽，它的变化速率就相当快，它与血纤维蛋白原结合在一起起着保护后者在血液凝结之前免于形成血纤维构型的功用，而一旦血液凝结，血纤维蛋白肽就在酶作用下被切除，放出血纤维蛋白原并形成血纤维，这样的功能似乎并不需要较高的结构特异性，因之亦相应倾向于较易产生变化。

通过各种生物血纤维蛋白肽一级顺序的比较，可估计出它的变化速率是每百万年改变该排列顺序的 1/100 残基，对球蛋白和细胞色素 *c* 来说，要改变 1/100 残基，则分别需要 600 万年及 2000 万年 (图 218)。于是人们会提出这样的问题，此种变化速率的差别是不是就真正反映了不同的蛋白质多肽分子所赋有的不同突变速率呢？还是仅仅因为细胞色素 *c* 基因中一个突变造成的危害是血纤维蛋白基因一个突变的 20 倍而产生的一种表面现象？(指二者突变速率相同但细胞色素 *c* 的突变保存率仅为 1/20。——译注)对这个问题有各种各样的回答，我们的倾向性看法是：细胞色素 *c* 分子由于有其高度专一性的功能部位以及它的兼有能同时与细胞色素还原酶及细胞色素氧化酶结合的特殊的结构状态，稍有变异即会产生有害的影响，因而在进化历史中表现为倾向保守，然而在细胞色素 *c* 分子中也有几个氨基酸残基可以自由变化而不致在功能上造成错误或失活，细胞色素 *c* 的这种功能对于专性需氧生物的生存是绝对必需的，因之，这一切便构成了决定某些氨基酸残基是否容忍其发生变异的限制因素。

下面我们将讨论的组蛋白Ⅳ有可能会使这一问题得到进一步的阐明。组蛋白Ⅳ有 102 个残基，在真核生物的染色体中，它和 DNA 结合在一起因而无所不在。对组蛋白Ⅳ初级结构顺序的比较研究表明，原来组蛋白Ⅳ是一类比细胞色素 *c* 还更加保守的蛋白质分子，实验估算差不多要经过 6 万万年才能改变其中

生物个体与大分子化合物进化的大致速率(从发生100个氨基酸残基变异所经历的地质时间估算)

马(生物个体的进化)		经历时间，年	组蛋白Ⅳ	细胞色素 *c*	球蛋白(Hb)	血纤维蛋白肽
		100,000				
		200,000				
		300,000				
		400,000				
		500,000			0.1	
	更新期	1百万年		0.1		1
马(属)					1	
			0.01			
	上新期	10百万年				10
				1		
上新马			0.01		10	
	始新期	100百万年				100
始祖马				10		
			1		100	
		10亿年				
				100		
			10			
		100亿年				

(a)

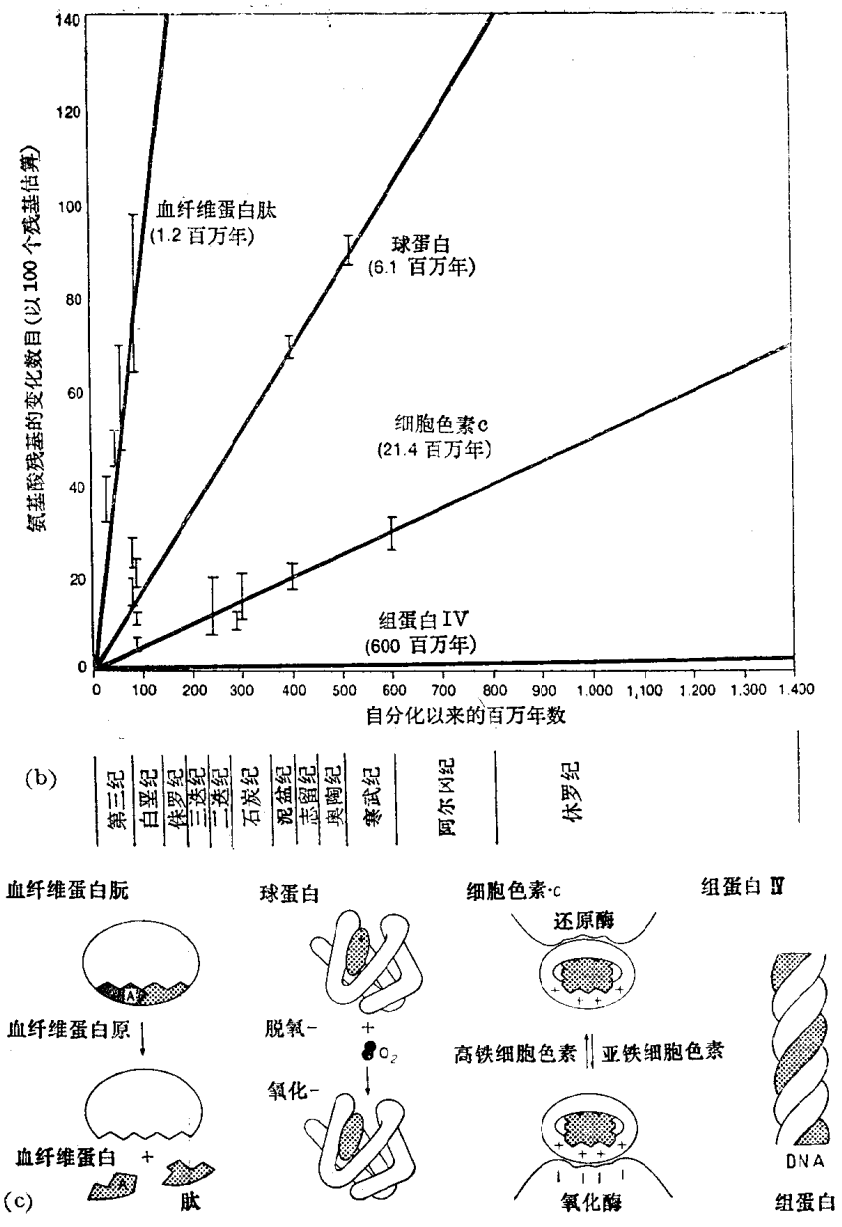


图 218　四种蛋白质的进化速率变化

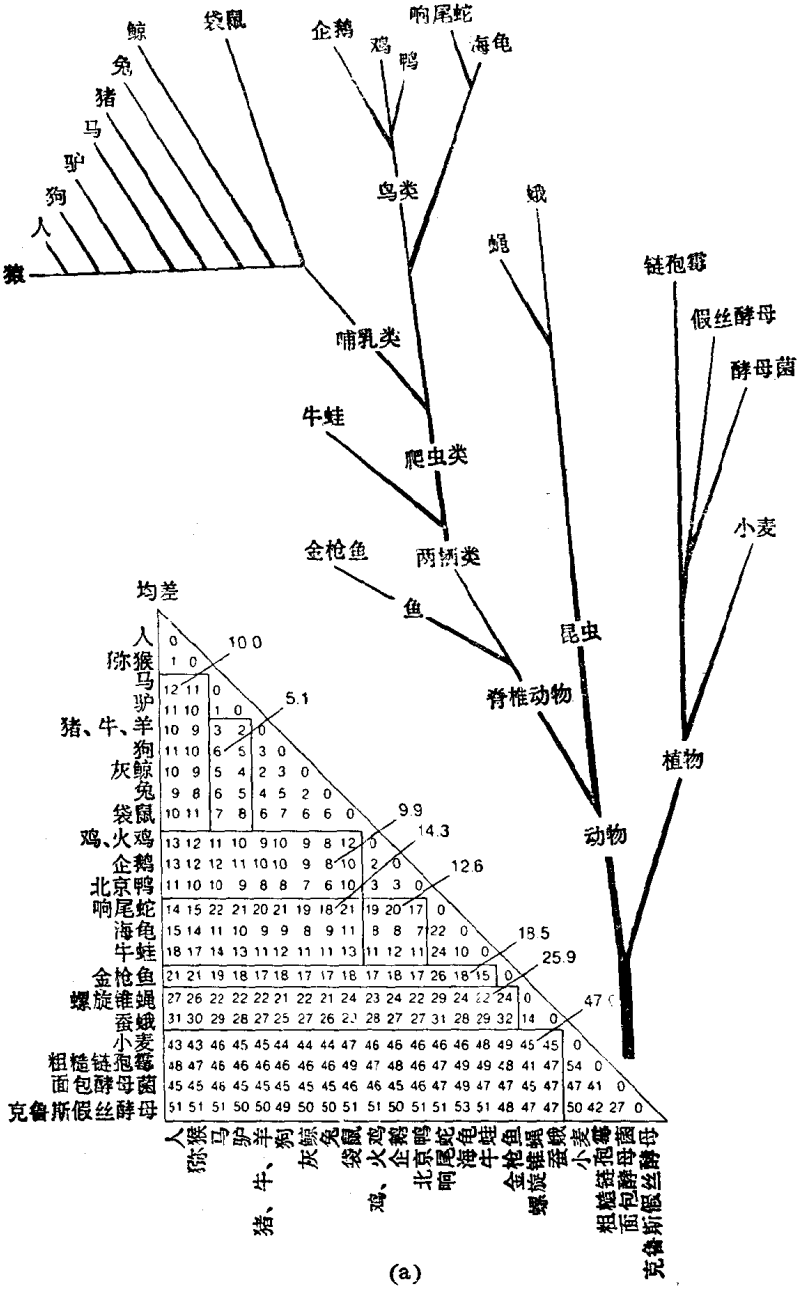
从现存动物中各种蛋白质类型一级结构上的差异所估算的进化速度，并将之与所经历的不同的地质时期进行对比(a 与 b)，一般公认限制蛋白质变异速度的因素是蛋白质的功能(c)，而不是基因本身的突变率，当然这样说并不意味着在不同种属的蛋白质分子中，每个密码子与每个基因的突变率一定都彼此等同。(引自 Dickerson 和 Geis, 1969, 及 Dickerson, 1972。)

1/100 的氨基酸顺序，这个速度是细胞色素 c 的 1/30。组蛋白所显示的这种高度的保守性，由于对它的功能还不全然明白，因而很难作出合理解释，据推测它的主要功能是调节 DNA 复制。

用蛋白质结构的多样性来解释进化关系在理论上还有许多漏洞，因为：(1)我们还无法判断这种从现代物种的蛋白质去探究已经消逝物种蛋白质的方法究竟有多大程度的可靠性，(2)还不可能估价在产生现代蛋白质分子的进化过程中，突变速率和选择因素各自所起的作用；(3)现在刚刚开始意识到"中性"突变的作用，尚无法评断"中性"突变究竟是如何地(如果有的话)影响着细胞色素 c 的进化。目前已知在一个物种内部并不存在互不相同的细胞色素 c 分子，从这一观察结果，有人猜想，即使确实有"中性"突变的话，恐怕也只会有极少数细胞色素 c 基因发生此类"中性"突变。

尽管现今解释蛋白质进化的理论有不少弱点，但是仍有它的可取之处。假如说建立进化的总体模式以及勾画现代物种与绝迹物种之间的系谱关系主要是依靠化石证据、比较解剖学、分类学、生理学等等方面的知识，那么蛋白质结构上的分歧程度同样也能够成为一种尺度，来测量那些最终会导致物种进化分歧的所可能发生的自然突变的数量。极端保守的蛋白质（如组蛋白Ⅳ——译注）正好显示出所有物种或差不多是所有物种之间的某种亲缘关系(有着同一的起源——译注)，而极易变化的蛋白质所说明的也正好是相对的少数几个物种之间的亲缘关系（如在结构上表现某几个方向的变化，则属于同一变化方向的物种亲缘关系较近，属于不同变化方向的物种亲缘关系较远——译注)。对上述这两种极端情况以及介于两者之间的各类蛋白质所作的分析表明，所有现存的物种在谱系关系上都可以将之归纳为一个统一的连续系列。既然在理论上可以凭借蛋白质的氨基酸排列顺序来追寻一个或几个中心起源的生命到所有现代物种的演化过程，因此全部生命进化史中的所谓知识空白以及失去的环节将会有可能被逐渐充实和弥合(图 219)。

蛋白质的顺序比较研究还揭示了具有不同功能的蛋白质分子


鲸
袋鼠
企鹅
响尾蛇
兔
鸡
鸭
海龟
猪
马
驴
狗
人
猿
鸟类
蛾
蝇
链孢霉
假丝酵母
哺乳类
酵母菌
牛蛙
爬虫类
小麦
金枪鱼
两栖类
鱼
昆虫
脊椎动物
植物
动物
均差
人
猕猴
马
驴
猪、牛、羊
狗
灰鲸
兔
袋鼠
鸡、火鸡
企鹅
北京鸭
响尾蛇
海龟
牛蛙
金枪鱼
螺旋锥蝇
蚕蛾
小麦
粗糙链孢霉
面包酵母菌
克鲁斯假丝酵母
100
5.1
9.9
14.3
12.6
18.5
25.9
47
(a)

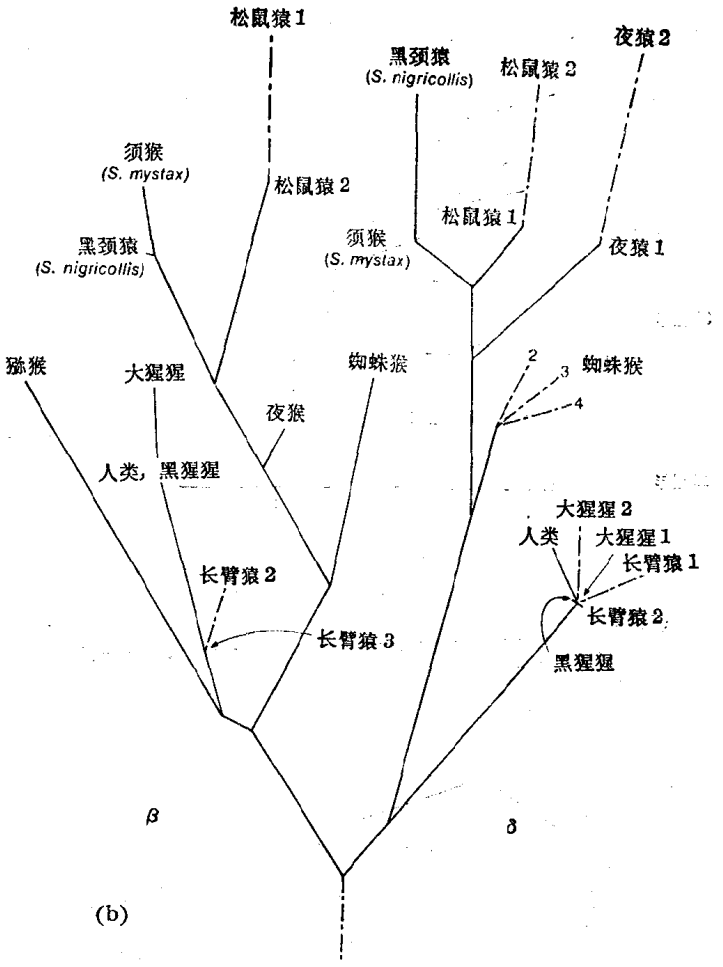
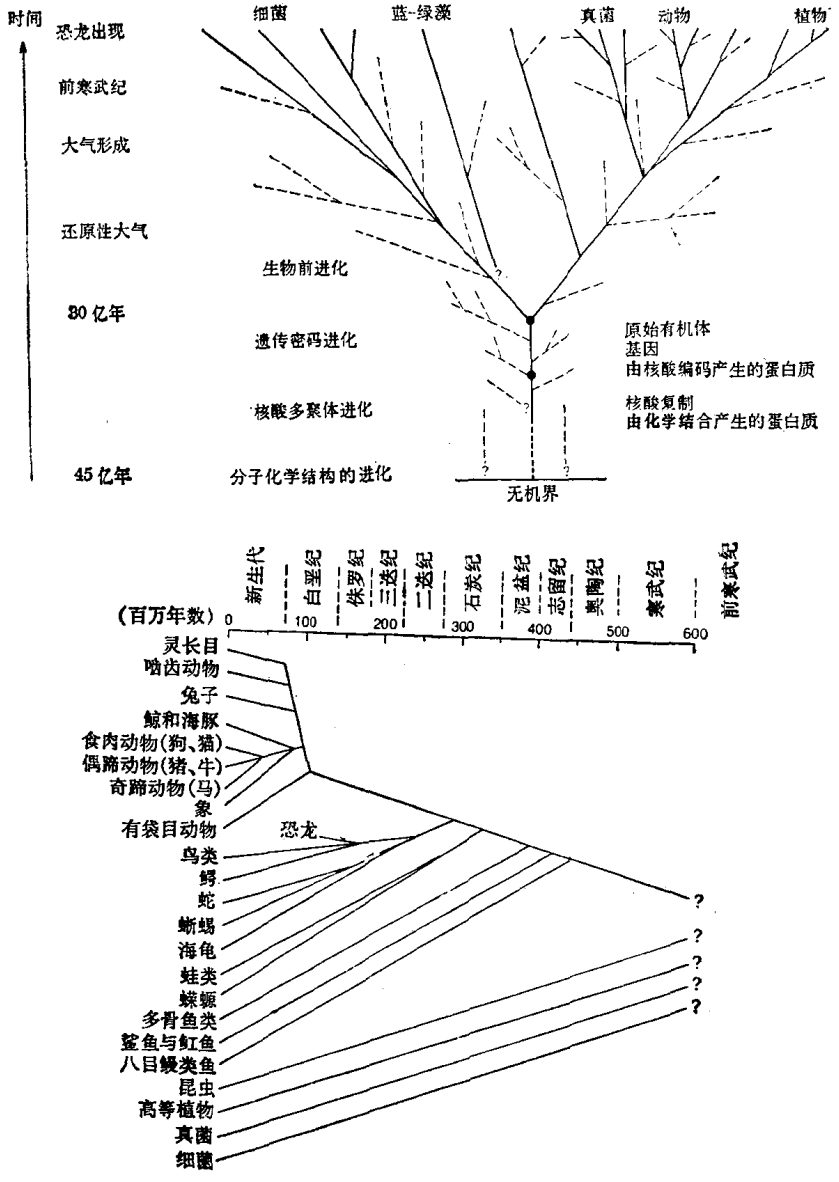


图 219　绘制进化系谱树的几种方法

(a) 许多用以建立系谱树的标准之所以可以信赖，是在于它们较客观地展示了不同物种之间的差异程度。至于"是怎么样的差别"肯定是不必说的。可是在分子水平上研究进化就很不一样。例如，从细胞色素c一级结构的比较研究表明(参见下图)，酵母菌和链孢霉这两个真菌间氨基酸的差异要比昆虫同脊椎动物之间的差异为大，在这里究竟"是怎样的差异"，数据是明白无误的。根据这一点无法解释这一矛盾，可是随着分析进一步深入，观察的数据愈来愈多，就会发现，从细胞色素c分子水平的差异也能估算有关种属的亲缘系谱，其中动植物所占的位置基本与其他标准的系谱关系大体符合(仅少数例外)。(引自 Dickerson 与 Geis, 1969。)

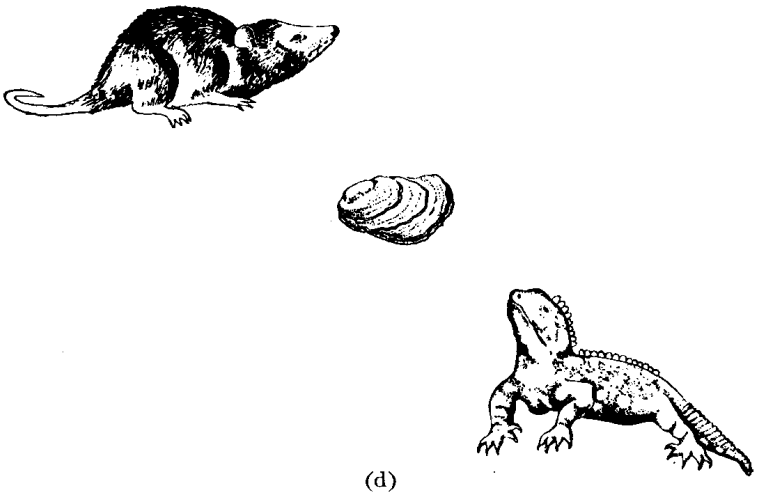
(b) 在比较血红蛋白β和δ链初级结构基础上绘成的谱系树，图中，连线长短是一种估算尺度，表示在彼此相互隔裂的动物类型之间可能有过的过渡类型突变型的数量。(引自 M. O. Dayhoff, 1972。)



(c)

图 219

(c) 根据地质年代(上)和化石证据(下)绘制的谱系树，上图虚线表示已经湮没无存的物种，下面"?"号则表示祖先种未定。(引自 M. O. Dayhoff, 1972。)



(d)

图 219

(d) 在一个极长时期内其形态基本没有什么改变的动物例证：鼠(上图)：8000万年；牡蛎：350 百万年；楔齿蜥：150 百万年。(引自 Simpson, 1949。)

之间的相互关系。例如具有 124 个氨基酸的 α-乳清蛋白的多肽链，其中有 53 个残基与鸡蛋白溶菌酶完全相同。两个蛋白质分子都有四个二硫键，所在的位置亦如出一辙，至于其他许多残基所表现的变异则是可以用非活性位点上的残基发生替换来解释的(图 220)。α-乳清蛋白并非是酶，它只是哺乳动物哺乳期时存在于乳腺中的乳糖合成酶这个复合物的一个组成部分。从历史的观点看问题，泌乳系统代谢进程的进化很可以说明 α-乳清蛋白和溶菌酶两者之间的关系。如果按照 Horowitz 的假说，即认为基因突变会使蛋白质的原有功能发生改变来解释酶的进化，我们就会理解，或许在最初的产乳系统中就一直是需要多糖裂解酶的，然后再从这个系统的多糖裂解酶最终又演化出了溶菌酶。

从前面就各种蛋白质类型所作的分析还很难说对于我们进一步认识进化过程作出了多大程度的贡献，到目前为止，这类研究只是从另一个侧面对进化的大致轮廓作了一番勾描，并澄清了某些有关种间进化关系上的模糊不清的概念，但是，更加重要的是，蛋白质初级结构的比较研究在认识进化的速率变化方面为人们提供

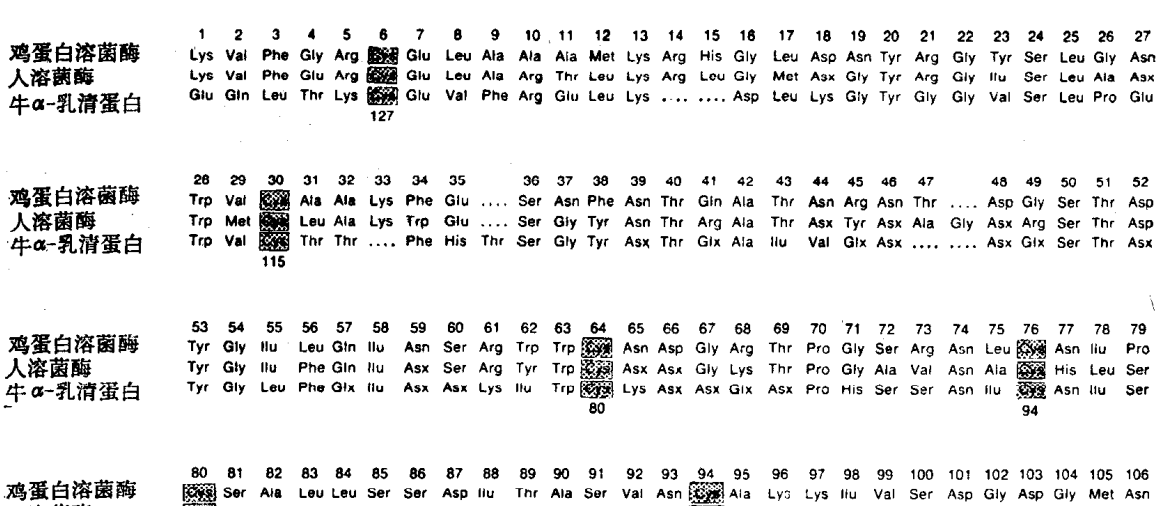


图 220　蛋白质分子功能不同但其一级顺序相似

溶菌酶与 α-乳清蛋白顺序中有许多残基是完全一样的，Glx 及 Asx 表示可能为谷氨酸，亦可能为谷氨酰胺，天冬氨酸亦可能为天冬酰胺，虚线表示有一个肽链分子顺序中有缺挡，也可能另一个肽链顺序中有插入。在判定顺序中位点的同源性时(指残基相同)如果把Glx及Asx视为 Gln、Asn便可使每一位点处于同源(如位点 59、121 等)，则定之为Gln 或 Asn。图中数字为鸡溶菌酶一级结构上的残基位点，在半胱氨酸残基下面标定的数字(共 8 处) 表示鸡溶菌酶及 α-乳清蛋白的四个二硫键的搭桥残基。

(引自 M. O. Dayhoff, 1972。)

了新的思路。看来，DNA 的突变率并不是进化速率的限制因素，现有证据表明，正是蛋白质结构和功能的制约关系影响着速率的变化。诚然，还可能有别的什么限制因子，但在结束讨论这个问题之前，还有必要指出，最近从有关功能上相关的酶(即同工酶)的三级结构比较研究已经进一步加强了这一论点。

事实上，从对蛋白质一级及三级结构所作的研究证据暗示，初级结构上显示的变化并不象三级结构那样，可以作为证明进化持续发展的可靠指标，例如，当进化的趋异性加强时，亦即在许多基因内部由于复制中发生差错而产生越来越多的突变时，与之相应的蛋白质分子的一级结构变异也越来越大，但在某些情况下(如变化的残基不处于关键位置)，纵然初级结构继续趋向于分歧多样，其三级结构亦未必受其影响，结果在功能上即表现为保守不变，从进化角度说，当然也谈不上发展。

目前，已经有几个酶，如乳酸脱氢酶(LDH)，可溶性苹果酸脱氢酶 (sMDH)、肝乙醇脱氢酶 (L-ADH)、3-磷酸甘油醛脱氢酶 (GPD)，在其三级结构上作了比较研究，结果发现，即使这些酶的一级结构、四级结构完全不同，但是却都有一个完全相同的大约由 60 个氨基酸组成的三级结构区。这个共同区域又包括两个结构区段，其中一个能使 2 个二聚体彼此聚合成四聚体，另一个则是与二核苷酸辅酶 (图中指 NAD) 结合的部位，图 221 展示的是 L-ADH 的三级结构及其与二核苷酸辅酶结合区段的图解。同样的区段也可以在别的脱氢酶中找到，当然这些酶分子的其余部分的氨基酸是彼此完全不同的。

二核苷酸辅酶参与电子及质子的传递在细胞代谢过程中具有极为重要的意义，因此不妨设想那些由二核苷酸辅酶协助作用的酶蛋白质在进化的早期历史阶段就已在行使它们的协同功能了。

这意味着，脱氢酶与诸如象 NAD 这样的二核苷酸辅酶相互间的结合作用，乃是一种一开始就必不可少的基本的生物学功能的需要，而维持这种特定的结合并不是取决于酶蛋白的初级结构，而是由脱氢酶三级结构所形成的某种特异性三度空间的构型区域

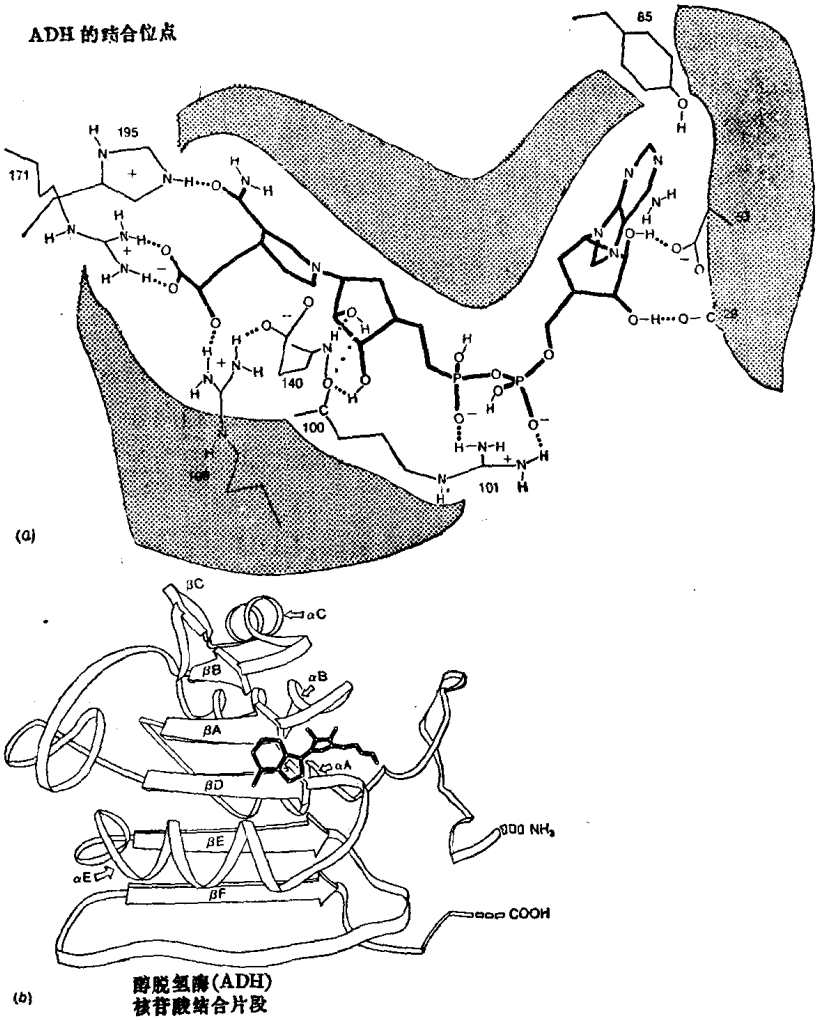


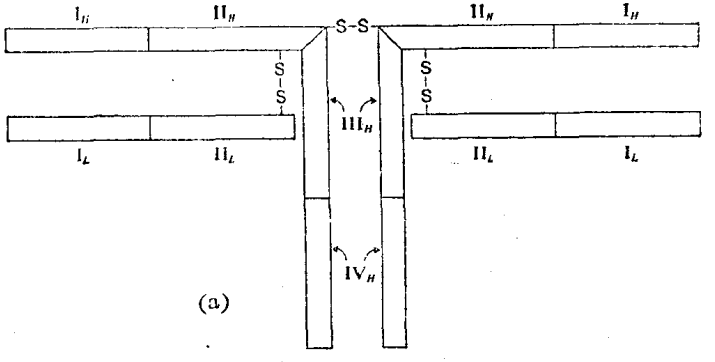
图 221　脱氢酶分子中与二核苷酸辅酶(NAD)的结合区段

(a) 图中所示为肽链中某些据以把底物 NAD 与结合区段连系在一起的活性部位的氨基酸残基，在其余脱氢酶中也可找到同样的区段。在脱氢酶系家族中，虽然结合区段的初级结构上有些变异，但其折迭方式完全相同，因之功能表现也不相上下。粗线表示二核苷酸辅酶，点线代表氢键，箭头为 β-折叠结构。

(b) 图中所示系为包括脱氢酶和二核苷酸辅酶的整个结合区段。(引自 A. Liljas 与 M. G. Rossman, 1974。)

来保证的。这类酶的其他区域可以变化，但一当由于某些氨基酸置换破坏了 6 个近似平行的β折叠结构和几个连串在一起的α螺旋所组成的区域的基本特征时，就会导致酶的失活。显然这样的氨基酸变换绝不会促进生物的进化发展过程。这样，当脱氢酶进化到具有利用种类繁多的底物催化多种反应的能力，并进而分化为彼此大不相同的酶蛋白时，所有这类酶的同二核苷酸辅酶相结合的特异性区域仍然会继承保留下来。从突变的角度来看，整个螺旋顺序有可能在蛋白质分子的某些部位造成缺损，而不使其功能丧失，可是单个氨基酸的置换只要它破坏了旨在保证酶分子发挥功能的基本折迭结构便往往是致死的。在生物种进化的漫长征途中，蛋白质的某些区域表现了高度保守，另一些区域中发生的氨基酸置换与缺失则会成为进化道路上分道扬镳的岔口，并由之发展出新的不同的生物种属。

上述概念亦为免疫球蛋白变异区的三维空间结构研究所证实。免疫球蛋白由 4 条多肽链组成，其中 2 条为重链，2 条为轻链(图 222a, b)，在重链和轻链的肽链上各有几个区段，肽链在此来回折迭，结构外形(似水桶壁)十分相似，桶壁区域共有大约 100 多个氨基酸残基，每条轻链有二个这样的区段，每条重链有四个这样的区段，每条肽链上的桶壁区段顺序实际上就是一个三级结构构型(图 222 c、d)。



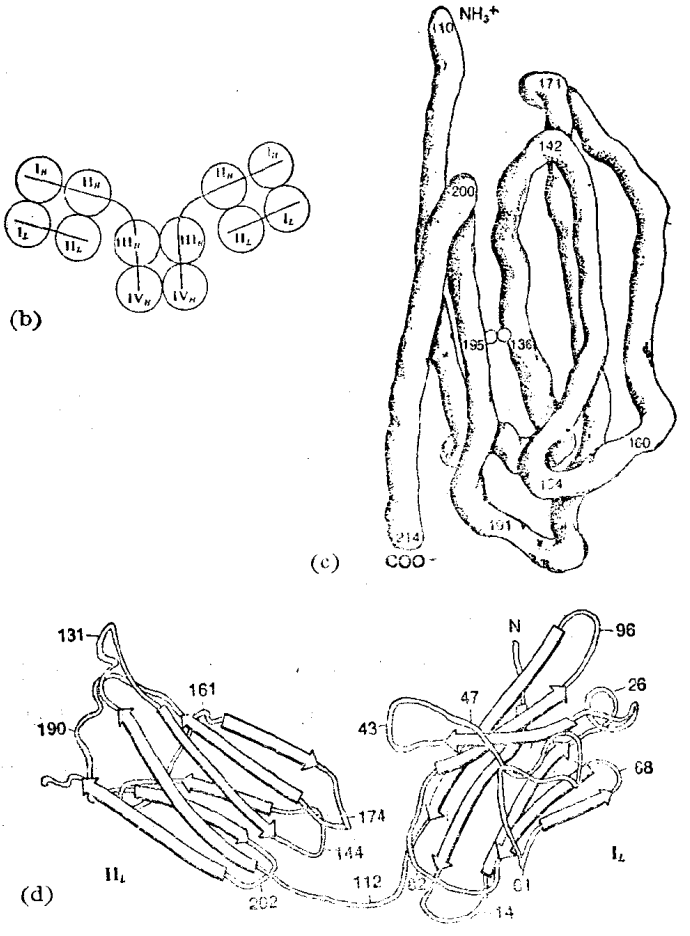


图 222 人类免疫球蛋白的三级结构和四级结构

(a) 免疫球蛋白的四级结构指的是四条多肽链包括两条重链和两条轻链的聚合构型，如图所示，这四条链部分地依靠二硫键(硫键桥)的作用连在一起，二种多肽链中都各包含几个大约拥有 100 个氨基酸残基组成的区域：重链有四个区(I_H, II_H, III_H, IV_H)，轻链有二个区(I_L, II_L)。

(b) 若用圆圈象征结构区域，即能清楚地看出四条链组成四级结构的相互关系，免疫球蛋白在I_H和I_L区段之间为抗原结合点。

(c) II_L结构区域的图解，免疫球蛋白分子链中的所有上述六个区段均与此相似，图中状似"桶"的"桶壁"部分为β-折迭结构。

(d) 图示一条轻链中二个状似桶壁的区段，数字表示肽链排列顺序中几个高度可变的氨基酸的大体位置，箭头代表β-折迭结构的走向。人们可以从这幅图中得到这样的印象，可变氨基酸组成区是由不变氨基酸的组成区规定和维持的。(引自 Schiffer 等, Biochem. 12:4620, 1973。)

在桶壁区段的氨基酸顺序中，有些位置上的氨基酸是高度可变的，而其他的氨基酸则是完全恒定的。对免疫球蛋白分子所进行的三级结构X-光衍射分析表明，可变区的顺序乃是那些很大程度上与维持桶壁状的三度空间折迭结构无关的区域，这暗示，凡是可能会带来基本空间结构毁坏的任何氨基酸置换都不会被保留下来。

以上列举的种种类型的蛋白结构比较研究都是一些新获得的材料，很难预见它们将来的进展情况，然而，有一点是清楚的，即分子生物学的技术和概念能作为分析工具和手段直接应用于研究进化过程。蛋白质结构的比较研究已经证明，在有些蛋白质中其十分重要的功能单位对于整个动植物界都是绝对不可缺少的，而在另一些蛋白质中，则不存在此种情况，在有亲缘关系的物种之间能够找到它们各种广泛变异的类型。任何对分子遗传及进化感兴趣的科学家当他们按照各个酶族自身的生物进化史来观察这些形形色色的酶分子的物理化学特性时，自必将会乐意遵循这一研究轨道继续前进。

参 考 文 献

Calvin, M.: *Chemical Evolution: Molecular Evolution toward the Origin of Living Systems on the Earth and Elsewhere*, Oxford University Press, London, 1969.

Fox, S. W., and K. Dose: *Molecular Evolution and the Origin of Life*, Freeman, San Francisco, 1972.

Margulis, L.: *Origin of Eucaryotic Cells*, Yale University Press, New Haven, 1970.

Mills, D. R., F. R. Kramer, and S. Spiegelman: Selfreplicating RNA, *Science*, 180:916 (1973).

Oparin, A. I.: The Origin of Life and the Origin of Enzymes, *Adv. Enzymol.*, 27:347 (1965).

Orgel, L. E.: Evolution of the Genetic Apparatus, *J. Mol. Biol.*, 38:381 (1968).

Rutten, M. G.: *The Origin of Life*, Elsevier, Amsterdam, 1971.

Woese, C. R.: *The Genetic Code*, Harper & Row, New York, 1967.

Stanbury, J. B., J. B. Wyngaarden, and D. B. Frederickson (eds.): *The Metabolic Basis of Inherited Disease*, 2d ed. McGraw-Hill, New York, 1966.

P21585